STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 111121

TO: Minh-Tam Davis **Location: cm1/8a01/8e12** **Art Unit: 1642** **Friday, January 02, 2004** **Case Serial Number: 09/554945**	**From: Mary Jane Ruhl** **Location: Biotech-Chem Library** **CM1-6A06** **Phone: 605-1155** **maryjane.ruhl@uspto.gov**

Search Notes

Examiner Davis,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
CM-1, Rm. 6-A-06
605-1155

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:20 ; Search time 46 Seconds
(without alignments)
1614.870 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2399	100.0	468	21	AAB18926	A novel polypeptid
2	2399	100.0	468	22	AAU12216	Human PRO5990 poly
3	2399	100.0	468	22	AAB87600	Human PRO5990. Ho
4	2399	100.0	468	23	ABG95925	Human secreted/tra
5	2399	100.0	468	24	ABU69097	Human PRO polypept
6	2399	100.0	468	24	ABU69120	Human PRO polypept
7	2399	100.0	468	24	ABU71580	Human secreted pol
8	2399	100.0	468	24	ABU72026	Novel human secret
9	2399	100.0	468	24	ABU72183	Human PRO polypept
10	2399	100.0	468	24	ABU66614	Human PRO polypept
11	2399	100.0	468	24	ABU66890	Human secreted/tra
12	2399	100.0	468	24	ABU59695	Novel secreted and
13	2399	100.0	468	24	ABG73287	Human neurosecreto
14	2399	100.0	567	22	AAO13905	Human polypeptide
15	2392	99.7	468	22	AAM93910	Human polypeptide,
16	2384	99.4	468	21	AAY81955	Human HsgIII prote
17	2271	94.7	470	22	ABG12976	Novel human diagno
18	2102.5	87.6	471	23	ABB57046	Mouse ischaemic co
19	1974	82.3	396	24	AAO16322	Human secreted pro
20	594	24.8	316	22	ABG14006	Novel human diagno
21	475	19.8	212	22	ABG14003	Novel human diagno
22	184	7.7	665	21	AAB18278	Plasmodium falcipa
23	180.5	7.5	1881	23	ABP73809	Candida albicans e
24	179	7.5	1558	21	AAB18324	Plasmodium falcipa
25	179	7.5	1786	18	AAW24790	P. falciparum live
26	179	7.5	1787	23	AAU96699	Plasmodium falcipo
27	178.5	7.4	1392	20	AAY06999	Restin protein seq
28	175.5	7.3	1427	12	AAR10534	Human 160kD mediat
29	174.5	7.3	1164	19	AAW40537	Group B streptococ
30	173	7.2	1099	19	AAW40538	Mutant C-beta prot
31	172.5	7.2	1164	17	AAR85781	Group B Streptococ
32	172.5	7.2	1164	21	AAY84459	Amino acid sequenc
33	172.5	7.2	1979	21	AAB18171	Plasmodium falcipa
34	171	7.1	1093	19	AAW40540	Mutant C-beta prot
35	170	7.1	1213	22	AAM40016	Human polypeptide
36	169.5	7.1	1788	22	ABB62995	Drosophila melanog
37	169	7.0	1129	21	AAY84462	Amino acid sequenc
38	168	7.0	1164	19	AAW40541	Mutant C-beta prot
39	168	7.0	1164	21	AAY84463	Amino acid sequenc
40	167	7.0	493	13	AAR26944	P.falciparum LSA g
41	166.5	6.9	1135	21	AAY84460	Amino acid sequenc
42	166	6.9	1128	19	AAW40539	Mutant C-beta prot
43	166	6.9	1128	21	AAY84461	Amino acid sequenc
44	166	6.9	1192	21	AAB18165	Plasmodium falcipa
45	166	6.9	2633	22	ABG06505	Novel human diagno

ALIGNMENTS

RESULT 1
AAB18926
ID AAB18926 standard; Protein; 468 AA.
XX
AC AAB18926;
XX
DT 08-FEB-2001 (first entry)
XX
DE A novel polypeptide designated PRO5990.
XX
KW Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW insulinemia; kidney disorder; Bergers disease; nephropathy;
KW Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW Crohns disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 68..71
FT /note= "N-glycosylation site"
FT Modified-site 70..73
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 82..85
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 97..100
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 125..128
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 143..148

```
FT                   /note= "N-myristoylation site"
FT   Modified-site   147..150
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   188..191
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   217..220
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   239..244
FT                   /note= "N-myristoylation site"
FT   Modified-site   265..268
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   289..292
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   305..308
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   320..323
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   326..329
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   346..349
FT                   /note= "N-glycosylation site"
FT   Modified-site   350..353
FT                   /note= "N-glycosylation site"
FT   Modified-site   362..365
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   369..372
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   382..385
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   386..389
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   387..390
FT                   /note= "casein kinase II phosphorylation site"
XX
PN   WO200056889-A2.
XX
PD   28-SEP-2000.
XX
PF   01-MAR-2000; 2000WO-US05601.
XX
PR   23-MAR-1999;   99US-0125774.
PR   23-MAR-1999;   99US-0125778.
PR   24-MAR-1999;   99US-0125826.
PR   31-MAR-1999;   99US-0127035.
PR   05-APR-1999;   99US-0127706.
PR   21-APR-1999;   99US-0130359.
PR   27-APR-1999;   99US-0131270.
PR   27-APR-1999;   99US-0131272.
PR   27-APR-1999;   99US-0131291.
PR   04-MAY-1999;   99US-0132371.
PR   04-MAY-1999;   99US-0132379.
PR   04-MAY-1999;   99US-0132383.
PR   25-MAY-1999;   99US-0135750.
PR   08-JUN-1999;   99US-0138166.
PR   20-JUL-1999;   99US-0144791.
PR   03-AUG-1999;   99US-0146970.
PR   09-DEC-1999;   99US-0170262.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
PI   Stewart TA, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2000-628263/60.
DR   N-PSDB; AAA96353.
XX
PT   Novel secreted and transmembrane polypeptides useful for diagnosing
PT   tumour in a mammal, for identifying agonists and antagonists of the
PT   polypeptide and for therapeutic use -
XX
PS   Claim 12; Fig 36; 222pp; English.
XX
CC   The present sequence represents a secreted or transmembrane polypeptide.
CC   The specification describes polypeptides designated PRO1484, PRO4334,
CC   PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC   PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC   PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC   useful for diagnosing tumour in a mammal. The polypeptides, their
CC   agonists and antagonists are useful treating a condition associated with
CC   expression or activity of the polypeptide. Conditions treated include
CC   obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are
CC   capable of inducing proliferation of mammalian kidney mesangial cells
CC   and are therefore useful for treating kidney disorders associated with
CC   decreased mesangial cell function such as Bergers disease or other
CC   nephropathies associated with Schonlein-Henoch purpura, celiac disease,
CC   dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
CC   to generate transgenic animals for use in development and screening of
CC   therapeutically useful reagents and also for chromosome identification
CC   and tissue typing.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 21;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 2
AAU12216
ID   AAU12216 standard; Protein; 468 AA.
XX
AC   AAU12216;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO5990 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI   Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI   Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21288.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical -
XX
PS   Claim 12; Fig 90; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   468 AA;
```

```
 Query Match           100.0%;  Score 2399;  DB 22;  Length 468;
 Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
 Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

```
RESULT 3
AAB87600
ID   AAB87600 standard; Protein; 468 AA.
XX
AC   AAB87600;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO5990.
XX
KW   Human; PRO protein; mapping.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
```

```
PI   Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI   Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92132.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 150; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the PRO protein,
CC   agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC   employed as molecular weight markers for protein electrophoresis. The PRO
CC   coding sequence has applications in molecular biology, including use as
CC   hybridisation probes, and in chromosome and gene mapping.
XX
SQ   Sequence   468 AA;

  Query Match           100.0%;  Score 2399;  DB 22;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 4
ABG95925
ID   ABG95925 standard; Protein; 468 AA.
XX
AC   ABG95925;
XX
DT   10-DEC-2002  (first entry)
XX
DE   Human secreted/transmembrane protein PRO5990.
XX
KW   Human; secreted protein; transmembrane protein; antirheumatic;
KW   antiarthritic; osteopathic; sports-related joint problem;
KW   articular cartilage defect; osteoarthritis; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
PN   US2002119130-A1.
XX
PD   29-AUG-2002.
XX
PF   06-DEC-2001; 2001US-0006867.
XX
PR   29-OCT-1997;   97US-063435P.
PR   29-OCT-1997;   97US-064215P.
PR   22-APR-1998;   98US-082797P.
PR   29-APR-1998;   98US-083495P.
PR   15-MAY-1998;   98US-085579P.
PR   10-JUN-1998;   98US-088811P.
PR   10-JUN-1998;   98US-088824P.
PR   10-JUN-1998;   98US-088825P.
PR   11-JUN-1998;   98US-088863P.
PR   12-JUN-1998;   98US-089105P.
PR   16-JUN-1998;   98US-089514P.
PR   16-SEP-1998;   98WO-US19330.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21194.
PR   22-DEC-1999;   99WO-US30720.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   30-MAR-2000; 2000WO-US08439.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   10-NOV-2000; 2000WO-US30873.
PR   01-DEC-2000; 2000WO-US32378.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI   Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX
DR   WPI; 2002-731348/79.
DR   N-PSDB; ABS74452.
XX
PT   New isolated secreted and transmembrane PRO polypeptide useful for
PT   modulating biological activity of a cell, or for treating
PT   sports-related joint problems, osteoarthritis or rheumatoid arthritis
XX
PS   Claim 20; Fig 150; 399pp; English.
XX
CC   The invention relates to an isolated secreted and transmembrane PRO
CC   polypeptide having 80 % sequence identity to a sequence appearing
CC   as ABG95851-ABG95934 or their associated signal peptide, or a sequence of
CC   an extracellular domain of the proteins with their associated signal
CC   peptide or lacking its associated signal peptide. Also included are
CC   the nucleic acids encoding the proteins, vectors, host cells,
CC   fusion proteins and antibodies which specifically bind to the proteins.
CC   The proteins are useful for detecting a polypeptide designated as A, B, C
CC   or D in a sample suspected of containing an A, B, C or D polypeptide,
CC   by contacting the sample with a polypeptide designated as E, F, G, H or
CC   I (or vice versa) and determining the formation of a A/E, B/F, B/G, C/H
CC   or D/I polypeptide conjugate in the sample, where the formation of the
CC   conjugate is indicative of the presence of an A, B, C or D polypeptide
CC   in the sample, where A is a PRO10272 polypeptide, B is a PRO20110
CC   polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,
CC   E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040
```

```
CC   polypeptide, H is a PRO20233 polypeptide and I is a PRO1890
CC   polypeptide. The sample comprises a cell suspected of expressing the A,
CC   B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with
CC   a detectable label or is attached to a solid support. The proteins are
CC   useful for linking a bioactive molecule to a cell expressing a
CC   polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive
CC   molecule is a toxin, a radiolabel or an antibody. The bioactive molecule
CC   causes death of the cell.  A, B, C, D, E, F, G, H, or I, or antibodies
CC   against them are useful for modulating a biological activity of a cell
CC   expressing a polypeptide designated as A, B, C or D or E, F, G, H, or
CC   I. The cell is killed. The proteins are useful for identifying
CC   agonists or antagonists, for the preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the proteins, as
CC   molecular weight markers for protein electrophoresis purposes, and as
CC   therapeutic agents for treating sports-related joint problems,
CC   articular cartilage defects, osteoarthritis or rheumatoid arthritis.
CC   Nucleic acids encoding the proteins are useful as hybridisation probes,
CC   in chromosome and gene mapping, in the generation of anti-sense RNA and
CC   DNA, for the preparation of the proteins, to generate transgenic or
CC   knockout animals which are useful in the development and screening of
CC   therapeutic useful reagents, for chromosome identification, and in gene
CC   therapy. The antibody is useful as a therapeutic agent, in a diagnostic
CC   assay and for affinity purification of the protein from recombinant
CC   cell culture natural sources. The present sequence represents a novel
CC   secreted or transmembrane protein of the invention.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 23;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 5
ABU69097
ID   ABU69097 standard; Protein; 468 AA.
XX
AC   ABU69097;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human PRO polypeptide #18.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein; cancer;
KW   non-insulin dependent diabetes mellitus; septic shock; stroke;
KW   rheumatoid arthritis; graft-versus-host disease; cardiac ischaemia;
KW   psoriasis; inflammatory bowel disease; asthma; antidiabetic;
KW   cytostatic; immunosuppressive; antirheumatic; antiarthritic;
KW   cerebroprotective; vasotropic; antipsoriatic; antiinflammatory;
KW   antiasthmatic.
XX
OS   Homo sapiens.
XX
PN   US2003008348-A1.
XX
PD   09-JAN-2003.
XX
PF   26-DEC-2001; 2001US-0035855.
XX
PR   14-MAY-1999;   99WO-US10733.
PR   02-DEC-1999;   99WO-US28551.
PR   22-DEC-1999;   99WO-US30720.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   15-MAY-1998;   98US-085579P.
PR   15-DEC-1998;   98US-112514P.
PR   22-DEC-1998;   98US-113300P.
PR   23-DEC-1998;   98US-113430P.
PR   23-DEC-1998;   98US-113605P.
PR   22-JAN-1999;   99US-116843P.
PR   23-MAR-1999;   99US-125774P.
PR   23-MAR-1999;   99US-125778P.
PR   24-MAR-1999;   99US-125826P.
PR   31-MAR-1999;   99US-127035P.
PR   05-APR-1999;   99US-127706P.
PR   13-APR-1999;   99US-129122P.
PR   21-APR-1999;   99US-130359P.
PR   25-MAY-1999;   99US-135750P.
PR   08-JUN-1999;   99US-138166P.
PR   20-JUL-1999;   99US-144791P.
PR   03-AUG-1999;   99US-146970P.
PR   29-OCT-1999;   99US-162506P.
PR   16-AUG-2001; 2001US-0931836.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2003-341326/32.
DR   N-PSDB; ACA06116.
XX
PT   New PRO polypeptides and nucleic acid molecules, useful for diagnosing
PT   or treating diabetes mellitus, cancers, septic shock, inflammatory
PT   bowel disease or asthma, or in gene therapy, chromosome identification
PT   or tissue typing  -
XX
PS   Claim 12; Fig 36; 196pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful in diagnosing or
CC   treating non-insulin dependent diabetes mellitus, cancers, septic
CC   shock, rheumatoid arthritis, graft-versus-host disease, stroke, cardiac
```

```
CC   ischaemia, psoriasis, inflammatory bowel disease or asthma. The PRO
CC   polynucleotide sequences may be used as hybridisation probes in
CC   chromosome and gene mapping, or in generating antisense RNA and DNA.
CC   They are also useful in preparing PRO polypeptides, in assays to
CC   identify other proteins or molecules involved in binding reaction, to
CC   generate transgenic animals or knockout animals, which in turn are
CC   useful in the development and screening of therapeutically useful
CC   reagents, for chromosome identification, and tissue typing. The PRO
CC   polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC   PRO polypeptides, or for the affinity purification of the polypeptides
CC   from recombinant cell culture or natural sources. ABU69080-ABU69102
CC   represent the human PRO polypeptides of the invention.
XX
SQ   Sequence   468 AA;

  Query Match           100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 6
ABU69120
ID   ABU69120 standard; Protein; 468 AA.
XX
AC   ABU69120;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human PRO polypeptide #18.
XX
KW   Human; secreted and transmembrane protein; bone disorder; obesity;
KW   cartilage disorder; sports injury; arthritis; diabetes mellitus;
KW   hypo-insulinaemia; obesity; hyper-insulinaemia; thalassaemia;
KW   haemoglobin-associated disorder; kidney disorder; Berger disease;
KW   mesangial cell function; nephropathy; Schonlein-Henoch purpura;
KW   celiac disease; dermatitis herpetiformis; Crohn's disease;
KW   anorectic; antiarthritic; antidiabetic; antianaemic; nephrotropic;
KW   antiinflammatory.
XX
OS   Homo sapiens.
XX
PN   US2003032061-A1.
XX
PD   13-FEB-2003.
XX
PF   26-DEC-2001; 2001US-0036214.
XX
PR   14-MAY-1999;   99WO-US10733.
PR   02-DEC-1999;   99WO-US28551.
PR   22-DEC-1999;   99WO-US30720.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   15-MAY-1998;   98US-085579P.
PR   15-DEC-1998;   98US-112514P.
PR   22-DEC-1998;   98US-113300P.
PR   23-DEC-1998;   98US-113430P.
PR   23-DEC-1998;   98US-113605P.
PR   23-DEC-1998;   98US-113621P.
PR   23-DEC-1998;   98US-114140P.
PR   12-JAN-1999;   99US-115552P.
PR   22-JAN-1999;   99US-116843P.
PR   23-MAR-1999;   99US-125774P.
PR   23-MAR-1999;   99US-125778P.
PR   24-MAR-1999;   99US-125826P.
PR   31-MAR-1999;   99US-127035P.
PR   05-APR-1999;   99US-127706P.
PR   13-APR-1999;   99US-129122P.
PR   21-APR-1999;   99US-130359P.
PR   27-APR-1999;   99US-131270P.
PR   27-APR-1999;   99US-131272P.
PR   27-APR-1999;   99US-131291P.
PR   04-MAY-1999;   99US-132371P.
PR   04-MAY-1999;   99US-132379P.
PR   04-MAY-1999;   99US-132383P.
PR   25-MAY-1999;   99US-135750P.
PR   08-JUN-1999;   99US-138166P.
PR   20-JUL-1999;   99US-144791P.
PR   03-AUG-1999;   99US-146970P.
PR   29-OCT-1999;   99US-162506P.
PR   16-AUG-2001; 2001US-0931836.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2003-341962/32.
DR   N-PSDB; ACA06173.
XX
PT   Novel isolated PRO polypeptides e.g., PRO4334, PRO1122, PRO1889,
PT   PRO1890, PRO1887, PRO1785, PRO4353, useful for treating sports
PT   injuries, arthritis, diabetes, obesity, hyper- or hypo-insulinaemia -
XX
PS   Claim 12; Fig 36; 194pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful in diagnosing or
CC   treating various bone and/or cartilage disorders (e.g. sports
CC   injuries, arthritis), various insulin deficient states (e.g. diabetes
```

```
CC   mellitus, hypo-insulinaemia), obesity, hyper-insulinaemia,
CC   haemoglobin-associated disorders (e.g. thalassaemias), kidney disorders
CC   associated with decreased mesangial cell function (e.g. Berger disease),
CC   or other nephropathies associated with Schonlein-Henoch purpura,
CC   celiac disease, dermatitis herpetiformis or Crohn's disease. The PRO
CC   polynucleotide sequences may be used as hybridisation probes in
CC   chromosome and gene mapping, or in generating antisense RNA and DNA.
CC   They are also useful in preparing PRO polypeptides, in assays to
CC   identify other proteins or molecules involved in binding reaction, to
CC   generate transgenic animals or knockout animals, which in turn are
CC   useful in the development and screening of therapeutically useful
CC   reagents, for chromosome identification, and tissue typing. The PRO
CC   polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC   PRO polypeptides, or for the affinity purification of the polypeptides
CC   from recombinant cell culture or natural sources. ABU69103-ABU69125
CC   represent the human PRO polypeptides of the invention.
XX
SQ   Sequence   468 AA;

  Query Match            100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 7
ABU71580
ID   ABU71580 standard; Protein; 468 AA.
XX
AC   ABU71580;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human secreted polypeptide PRO5990.
XX
KW   Human; gene therapy; tumour; cancer.
XX
OS   Homo sapiens.
XX
PN   US2003013855-A1.
XX
PD   16-JAN-2003.
XX
PF   03-MAY-2002; 2002US-0063616.
XX
PR   30-DEC-1998;   98KR-0062142.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   30-DEC-1999;   99WO-US31274.
PR   18-FEB-2000; 2000WO-US04341.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   21-MAR-2000; 2000WO-US07532.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   24-AUG-2000; 2000WO-US23328.
PR   10-NOV-2000; 2000WO-US30873.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   14-MAY-1999;   99US-0311832.
PR   25-AUG-1999;   99US-0380137.
PR   25-AUG-1999;   99US-0380138.
PR   25-AUG-1999;   99US-0380139.
PR   25-AUG-1999;   99US-0380142.
PR   15-SEP-1999;   99US-0397342.
PR   18-OCT-1999;   99US-0403297.
PR   12-NOV-1999;   99US-0423844.
PR   22-AUG-2000; 2000US-0644848.
PR   18-SEP-2000; 2000US-0664610.
PR   18-SEP-2000; 2000US-0665350.
PR   08-NOV-2000; 2000US-0709238.
PR   20-DEC-2000; 2000US-0747259.
PR   22-MAR-2001; 2001US-0816744.
PR   10-MAY-2001; 2001US-0854208.
PR   10-MAY-2001; 2001US-0854280.
PR   30-MAY-2001; 2001US-0870574.
PR   05-JUN-2001; 2001US-0874503.
PR   29-JUN-2001; 2001US-0869599.
PR   18-JUL-2001; 2001US-0908827.
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-330485/31.
DR   N-PSDB; ACA58884.
XX
PT   New isolated antibody specifically binding a PRO polypeptide, useful
PT   for the preparation of a medicament for treating disorders with the
PT   aberrant expression or activity of the PRO polypeptide, such as tumor
PT   conditions and cancer  -
XX
PS   Example 17; Page 218-219; 406pp; English.
XX
CC   The invention relates to an antibody that binds to a polypeptide with a
CC   fully defined sequence given in the specification. The methods and
CC   compositions (containing antibodies that specifically bind a PRO
CC   polypeptide) of the present invention are useful for the preparation of a
CC   medicament for the treatment of disorders associated with the aberrant
CC   expression or activity of the PRO polypeptide, such as tumour conditions
CC   and cancer. They can also be used to generate transgenic or knockout
CC   animals useful in the development and screening of therapeutically useful
CC   reagents. The PRO polypeptides and encoding nucleic acids can be used as
CC   molecular weight markers for protein electrophoresis, chromosome
CC   identification and tissue typing. The PRO polypeptides are useful to
CC   induce angiogenesis e.g wound healing; in the treatment of sports-related
CC   joint problems, articular cartilage defects, osteoarthritis or rheumatoid
CC   arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The
CC   antibodies may be used in various diagnostic, competitive binding and/or
```

```
CC   immunoprecipitation assays. The present sequence represents the amino
CC   acid sequence of a PRO polypeptide of the invention.
XX
SQ   Sequence   468 AA;

  Query Match           100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 8
ABU72026
ID   ABU72026 standard; Protein; 468 AA.
XX
AC   ABU72026;
XX
DT   11-JUN-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO5990.
XX
KW   Human; secreted and transmembrane polypeptide;
KW   chromosome mapping; gene mapping; transgenic animal; knockout animal;
KW   therapeutic agent screening; chromosome identification; tissue typing;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   US2003018183-A1.
XX
PD   23-JAN-2003.
XX
PF   01-MAY-2002; 2002US-0063512.
XX
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-330984/31.
DR   N-PSDB; ACA60437.
XX
PT   New secreted and transmembrane PRO polypeptides and nucleic acid
PT   molecules encoding the polypeptides, useful in gene therapy or
PT   preparing a medicament for treating a condition that is responsive to
PT   the PRO polypeptide or antibody -
XX
PS   Disclosure; Fig 150; 409pp; English.
XX
CC   The invention describes novel isolated PRO polypeptides. The PRO
CC   polypeptides or anti-PRO antibodies are useful in preparing a medicament
CC   for treating a condition that is responsive to the PRO polypeptide or
CC   antibody. The PRO nucleotide sequences may be used as hybridisation
CC   probes in chromosome and gene mapping, or in generating antisense RNA
CC   and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,
CC   in assays to identify other proteins or molecules involved in binding
CC   reaction, to generate transgenic animals or knockout animals, which in
CC   turn are useful in the development and screening of therapeutically
CC   useful reagents, for chromosome identification, and tissue typing. The
CC   PRO polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. The anti-PRO antibodies may be used in diagnostic assays for
CC   PRO, or for the affinity purification of PRO from recombinant cell
CC   culture or natural sources. This is the amino acid sequence of a novel
CC   human secreted and transmembrane PRO poypeptide.
XX
SQ   Sequence   468 AA;

  Query Match           100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 9
ABU72183
ID   ABU72183 standard; Protein; 468 AA.
XX
AC   ABU72183;
XX
DT   13-JUN-2003  (first entry)
XX
```

```
DE   Human PRO polypeptide #75.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein;
KW   anti-PRO antibody; diagnostic assay; gene expression.
XX
OS   Homo sapiens.
XX
PN   US2003023042-A1.
XX
PD   30-JAN-2003.
XX
PF   01-MAY-2002; 2002US-0063502.
XX
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-331484/31.
DR   N-PSDB; ACA63447.
XX
PT   Novel monoclonal antibody that binds to secreted and transmembrane
PT   polypeptide, useful for detecting and purifying the polypeptide and
PT   also for treating conditions responsive to the antibody  -
XX
PS   Disclosure; Fig 150; 408pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful for preparing a
CC   medicament useful in the treatment of a condition responsive to
CC   anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic
CC   assays for PRO, by detecting its expression in specific cells,
CC   tissues or serum, and for affinity purification of PRO from
CC   recombinant cell culture or natural sources. ABU72109-ABU72192
CC   represent the human PRO polypeptides of the invention.
XX
SQ   Sequence   468 AA;

  Query Match            100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 10
ABU66614
ID   ABU66614 standard; Protein; 468 AA.
XX
AC   ABU66614;
XX
DT   23-MAY-2003  (first entry)
XX
DE   Human PRO polypeptide #45.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein;
KW   tumour necrosis factor-alpha; TNF-alpha; blood; proliferation;
KW   differentiation; chondrocyte; tumour; genetic disorder;
KW   cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003036180-A1.
XX
PD   20-FEB-2003.
XX
PF   09-MAY-2002; 2002US-0143114.
XX
PR   31-MAR-1997;   97WO-US05230.
PR   12-JUN-1998;   98WO-US12456.
PR   14-JUL-1998;   98WO-US14552.
PR   28-AUG-1998;   98WO-US17888.
PR   10-SEP-1998;   98WO-US18824.
PR   14-SEP-1998;   98WO-US19093.
PR   14-SEP-1998;   98WO-US19094.
PR   14-SEP-1998;   98WO-US19177.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   29-OCT-1998;   98WO-US22991.
PR   29-OCT-1998;   98WO-US22992.
PR   20-NOV-1998;   98WO-US24855.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   10-MAR-1999;   99WO-US05190.
PR   20-APR-1999;   99WO-US08615.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   22-DEC-1999;   99WO-US30720.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
```

PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05746.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 25-MAY-2001; 2001WO-US17092.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 22-JUN-2001; 2001WO-US20116.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 20-DEC-2000; 2000US-0747259.
PR 28-FEB-2001; 2001US-0796498.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 18-MAY-2001; 2001US-0860216.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 01-JUN-2001; 2001US-0872035.
PR 05-JUN-2001; 2001US-0874503.
PR 14-JUN-2001; 2001US-0882636.
PR 19-JUN-2001; 2001US-0886342.
PR 21-JUN-2001; 2001US-0887879.
PR 18-JUL-2001; 2001US-0908827.
PR 06-AUG-2001; 2001US-0924419.
PR 09-AUG-2001; 2001US-0927796.
PR 16-AUG-2001; 2001US-0931836.
PR 19-DEC-2001; 2001US-0028072.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2003-332040/31.
DR N-PSDB; ACA03647.
XX
PT New secreted and transmembrane PRO nucleic acids, useful for gene
PT therapy, in chromosome and gene mapping, as chromosome markers, in
PT tissue typing, and in chromosome identification -
XX
PS Claim 12; Fig 90; 660pp; English.
XX
CC The present invention relates to the isolation of novel human PRO
CC polypeptides, and the polynucleotide sequences encoding them. The
CC PRO polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides are useful for detecting other PRO polypeptides, for
CC linking bioactive molecules to cells expressing PRO polypeptides,
CC for modulating biological activities of cells expressing PRO
CC polypeptides, and for for identifying agonists or antagonists.
CC The PRO polypeptides are useful for for stimulating the release of
CC tumour necrosis factor (TNF)-alpha from human blood, for stimulating
CC the proliferation or differentiation of chondrocytes, and detecting the
CC presence of tumours. The polynucleotide sequences encoding PRO
CC polypeptides are useful as hybridisation probes, in chromosome and
CC gene mapping, in the generation of antisense RNA and DNA, in the
CC preparation of PRO polypeptides, for generating transgenic animals or
CC knockout animals, for the genetic analysis of individuals with genetic
CC disorders, and in gene therapy. ABU66570-ABU66844 represent the human
CC PRO polypeptides of the invention.
CC Note: The sequence data for this patent was obtained in electronic
CC format directly from the USPTO web site at
CC seqdata.uspto.gov/psipsDIDEntry.html.
XX
SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 11
ABU66890
ID ABU66890 standard; Protein; 468 AA.
XX
AC ABU66890;
XX
DT 27-MAY-2003 (first entry)
XX
DE Human secreted/transmembrane, PRO, protein SEQ ID 90.
XX
KW Human; secreted protein; transmembrane protein; PRO;
KW inflammatory disease; organ failure; atherosclerosis; cardiac injury;
KW infertility; birth defects; premature aging; AIDS; biosensor;
KW acquired immunodeficiency syndrome; cancer; diabetic complication;
KW bioreactor; tumour.

XX
OS Homo sapiens.
XX
PN US2003032155-A1.
XX
PD 13-FEB-2003.
XX
PF 03-MAY-2002; 2002US-0137865.
XX
PR 31-MAR-1997; 97WO-US05230.
PR 12-JUN-1998; 98WO-US12456.
PR 14-JUL-1998; 98WO-US14552.
PR 28-AUG-1998; 98WO-US17888.
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19093.
PR 14-SEP-1998; 98WO-US19094.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 29-OCT-1998; 98WO-US22991.
PR 29-OCT-1998; 98WO-US22992.
PR 20-NOV-1998; 98WO-US24855.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 99WO-US00106.
PR 08-MAR-1999; 99WO-US05028.
PR 10-MAR-1999; 99WO-US05190.
PR 20-APR-1999; 99WO-US08615.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 22-DEC-1999; 99WO-US30720.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05746.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 25-MAY-2001; 2001WO-US17092.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 22-JUN-2001; 2001WO-US20116.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 20-DEC-2000; 2000US-0747259.
PR 28-FEB-2001; 2001US-0796498.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 18-MAY-2001; 2001US-0860216.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 01-JUN-2001; 2001US-0872035.
PR 05-JUN-2001; 2001US-0874503.
PR 14-JUN-2001; 2001US-0882636.
PR 19-JUN-2001; 2001US-0886342.
PR 21-JUN-2001; 2001US-0887879.
PR 18-JUL-2001; 2001US-0908827.
PR 06-AUG-2001; 2001US-0924419.
PR 09-AUG-2001; 2001US-0927796.
PR 16-AUG-2001; 2001US-0931836.
PR 19-DEC-2001; 2001US-0028072.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2003-331925/31.
DR N-PSDB; ACA04068.
XX
PT New secreted and transmembrane nucleic acids and polypeptides,
PT designated as PRO, useful for treating inflammation, organ failure,
PT atherosclerosis, cardiac injury, infertility, birth defects, premature
PT aging, AIDS, or cancer -
XX
PS Claim 12; Fig 90; 659pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising, or which is
CC at least 80% identical to, or the full-length coding sequence of, any of
CC the 275 nucleotide sequences, encoding the corresponding PRO polypeptide
CC (one of 275 secreted or transmembrane proteins). The nucleic acid
CC further comprises the full-length coding sequence of the DNA deposited
CC under American Type Culture Collection (ATCC) accession number in a list
CC given in the specification. Also included are vectors and host
CC cells for producing PRO proteins, PRO fusion proteins, anti-PRO
CC antibodies, PRO extracellular domains and mature sequences, methods
CC of detecting PRO proteins, methods for stimulating the release of
CC TNF-alpha (tumour necrosis factor alpha) from human blood,
CC (and the proliferation of differentiation of chondrocyte cells, the
CC proliferation of, or gene expression in pericyte cells, the release or
CC proteoglycans from cartilage, proliferation of inner ear urticular
CC supporting cells, the proliferation of T-lymphocyte cells, the release
CC of a cytokine from peripheral blood mononuclear cells (PBMC), or the
CC proliferation of endothelial cells), a method for modulating the uptake
CC of glucose or free fatty acid (FFA) by skeletal muscle cells,
CC a method for inhibiting the binding of A-peptide to factor VIIA,
CC or the differentiation of adipocyte cells, a method for detecting the
CC presence of a tumour in a mammal and an oligonucleotide probe derived

```
CC   from any of the nucleotide sequences cited above. The nucleic acids and
CC   polypeptides are useful for treating inflammatory diseases, organ
CC   failure, atherosclerosis, cardiac injury, infertility, birth defects,
CC   premature aging, AIDS (acquired immunodeficiency syndrome), cancer, or
CC   diabetic complications. The nucleic acids are useful as hybridisation
CC   probes, in chromosome and gene mapping, and in generating antisense RNA
CC   or DNA. The polypeptides are useful as pharmaceuticals, diagnostics,
CC   biosensors or bioreactors. Both are useful in tissue typing.
CC   The present sequence represents a PRO protein of the invention.
XX
SQ   Sequence   468 AA;

  Query Match            100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 12
ABU59695
ID   ABU59695 standard; Protein; 468 AA.
XX
AC   ABU59695;
XX
DT   13-MAY-2003  (first entry)
XX
DE   Novel secreted and transmembrane protein PRO5990.
XX
KW   Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;
KW   cardiac insufficiency disorder; cancer; tumour; immune response;
KW   adrenal cortical capillary endothelial growth; c-fos induction;
KW   vascular endothelial growth factor inhibition; VEGF inhibition;
KW   endothelial cell growth inhibitor; T-lymphocytes stimulation;
KW   retinal neurons cell survival; rod photoreceptor cell survival;
KW   retinal disorder; retinitis pigmentosum; kidney disorder;
KW   mammalian kidney mesangial cell proliferation; Berger disease;
KW   dermatitis; herpetiformis; Crohn's disease; chondrocyte proliferation;
KW   chondrocyte redifferentiation; sports injury; arthritis.
XX
OS   Homo sapiens.
XX
PN   US2003017563-A1.
XX
PD   23-JAN-2003.
XX
PF   07-MAY-2002; 2002US-0140808.
XX
PR   31-MAR-1997;   97WO-US05230.
PR   12-JUN-1998;   98WO-US12456.
PR   14-JUL-1998;   98WO-US14552.
PR   28-AUG-1998;   98WO-US17888.
PR   10-SEP-1998;   98WO-US18824.
PR   14-SEP-1998;   98WO-US19093.
PR   14-SEP-1998;   98WO-US19094.
PR   14-SEP-1998;   98WO-US19177.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   29-OCT-1998;   98WO-US22991.
PR   29-OCT-1998;   98WO-US22992.
PR   20-NOV-1998;   98WO-US24855.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   10-MAR-1999;   99WO-US05190.
PR   20-APR-1999;   99WO-US08615.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   22-DEC-1999;   99WO-US30720.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05746.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   11-AUG-2000; 2000WO-US22031.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   10-NOV-2000; 2000WO-US30873.
```

PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 25-MAY-2001; 2001WO-US17092.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 22-JUN-2001; 2001WO-US20116.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 20-DEC-2000; 2000US-0747259.
PR 28-FEB-2001; 2001US-0796498.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 18-MAY-2001; 2001US-0860216.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 01-JUN-2001; 2001US-0872035.
PR 05-JUN-2001; 2001US-0874503.
PR 14-JUN-2001; 2001US-0882636.
PR 19-JUN-2001; 2001US-0886342.
PR 21-JUN-2001; 2001US-0887879.
PR 18-JUL-2001; 2001US-0908827.
PR 06-AUG-2001; 2001US-0924419.
PR 09-AUG-2001; 2001US-0927796.
PR 16-AUG-2001; 2001US-0931836.
PR 19-DEC-2001; 2001US-0028072.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2003-148238/14.
DR N-PSDB; ABX89185.
XX
PT Novel isolated PRO polypeptides e.g., PRO826, PRO1068, PRO1184, PRO1346
PT and PRO1375, which stimulate proliferation of stimulated T-lymphocytes
PT are therapeutically useful for enhancing immune response and in cancer
PT treatments -
XX
PS Claim 12; Fig 90; 659pp; English.
XX
CC The invention describes an isolated human PRO polypeptide. The PRO
CC polypeptides are useful in detecting PRO polypeptides in a sample, in
CC linking a bioactive molecule to a cell expressing a PRO polypeptide, and
CC in modulating at least one biological activity of a cell expressing a PRO
CC polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus
CC useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186
CC stimulate adrenal cortical capillary endothelial growth, and PRO536,
CC PRO943, PRO828, PRO826, PRO1068 or PRO535, PRO826, PRO819, PRO1126,
CC PRO1360 and PRO1387 induce c-fos in endothelial cells, and are thus
CC useful for treating conditions or disorders where angiogenesis would be
CC beneficial, e.g. wound healing and antagonist of this polypeptide are
CC useful for treating cancerous tumours. PRO812 inhibits vascular
CC endothelial growth factor (VEGF) stimulated proliferation of endothelial
CC cells and is thus useful for inhibiting endothelial cell growth in
CC mammals which would be beneficial in inhibiting tumour growth. PRO826,
CC PRO1068, PRO1184, PRO1346 and PRO1375 stimulate proliferation of
CC stimulated T-lymphocytes and are therapeutically useful for enhancing
CC immune response. PRO828, PRO826, PRO1068 or PRO1132 enhance survival of
CC retinal neurons cells (PRO1132 is also enhances survival/proliferation of
CC rod photoreceptor cells) and therefore are useful for treating retinal
CC disorders of injuries, e.g. retinitis pigmentosum, AMD. PRO819, PRO813
CC and PRO11066 induce proliferation of mammalian kidney mesangial cells,
CC and therefore are useful for treating kidney disorders associated with
CC decreased mesangial cell function such as Berger disease or other
CC nephropathies associated with dermatitis, herpetiformis or Crohn's
CC disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the
CC proliferation and/or redifferentiation of chondrocytes in culture and
CC are thus useful for treating sports injuries, and arthritis. This
CC is the amino acid sequence of a novel human PRO protein.
XX
SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 13
ABG73287
ID ABG73287 standard; Protein; 468 AA.
XX
AC ABG73287;
XX
DT 28-APR-2003 (first entry)
XX
DE Human neurosecretory protein-1 (HUNSP-1).
XX
KW Human; neurosecretory protein-1; HUNSP-1; neurological disorder;
KW epilepsy; Huntington's disease; dementia; Alzheimer's disease;
KW multiple sclerosis; Parkinson's disease; schizophrenia; amnesia;
KW endocrine disorder; hypogonadism; Sheehan's syndrome; leukaemia;
KW Hashimoto's disease; Grave's disease; Cushing's disease; cancer;
KW adenocarcinoma; diabetes insipidus; melanoma; myeloma sarcoma;
KW bone cancer; breast cancer; cancer of the pancreas; brain cancer;
KW prostate cancer; skin cancer; thyroid cancer; antidiabetic; nootropic;
KW neuroprotective; neuroleptic; cytostatic; anticonvulsant; thyromimetic;
KW lymphoma; antiparkinsonian; antithyroid; immunosuppressive; antithyroid;
KW pancreatic islet cell cDNA library; ISLTNOT01; Incyte clone 2379427.
XX
OS Homo sapiens.
XX
PN US2002156014-A1.
XX
PD 24-OCT-2002.
XX
PF 14-FEB-2002; 2002US-0077040.

```
XX
PR   17-APR-1998;   98US-0062601.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P,  Corley NC,  Patterson C;
XX
DR   WPI; 2003-238117/23.
DR   N-PSDB; ABX11115.
XX
PT   New human neurosecretory proteins (HUNSP) and polynucleotides, useful
PT   for diagnosing, treating or preventing e.g. epilepsy, Alzheimer's
PT   disease, schizophrenia, Grave's disease, leukaemia, lymphoma, or breast
PT   or brain cancer  -
XX
PS   Claim 1; Fig 1; 45pp; English.
XX
CC   The present invention relates to the isolation of human neurosecretory
CC   proteins -1 and -2 (HUNSP-1 and HUNSP-2), and the polynucleotide
CC   sequences encoding them. The HUNSP polypeptides, polynucleotides, and
CC   agonists are useful for treating a disease or condition associated with
CC   decreased expression of functional HUNSP. Antagonists of HUNSP are
CC   useful for treating a disease or condition associated with
CC   overexpression of functional HUNSP. HUNSP antibodies are useful for
CC   diagnosing a condition or disease associated with the expression of
CC   HUNSP. The sequences of the invention are particularly useful for
CC   diagnosing, treating or preventing neurological disorders
CC   (e.g. epilepsy, Huntington's disease, dementia, Alzheimer's disease,
CC   multiple sclerosis, Parkinson's disease, schizophrenia, amnesia),
CC   endocrine disorders (e.g. hypogonadism, Sheehan's syndrome, diabetes
CC   insipidus, Hashimoto's disease, Grave's disease, Cushing's disease),
CC   and cancers (e.g. adenocarcinoma, leukaemia, lymphoma, melanoma,
CC   myeloma sarcoma, bone cancer, breast cancer, cancer of the pancreas,
CC   brain cancer, prostate cancer, skin cancer or thyroid cancer).
CC   The present sequence represents HUNSP-1. The cDNA encoding HUNSP-1
CC   is isolated from pancreatic islet cell cDNA library (ISLTNOT01) Incyte
CC   clone 2379427.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 14
AAO13905
ID   AAO13905 standard; Protein; 567 AA.
XX
AC   AAO13905;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 27797.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI93836.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 27797; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   567 AA;

  Query Match             100.0%;  Score 2399;  DB 22;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 159

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 219

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 279

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 339

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 399

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 459

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 519

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
                 ||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 567
```

RESULT 15
AAM93910
ID AAM93910 standard; Protein; 468 AA.
XX
AC AAM93910;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 4061.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94872.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 4061; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX
SQ Sequence 468 AA;

Query Match 99.7%; Score 2392; DB 22; Length 468;
Best Local Similarity 99.8%; Pred. No. 1.2e-145;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEGIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

Search completed: January 2, 2004, 12:34:22
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 21 Seconds
(without alignments)
942.928 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	179	7.5	1786	3	US-08-973-462-8	Sequence 8, Appli
2	172.5	7.2	1164	3	US-08-923-992A-2	Sequence 2, Appli
3	169	7.0	1098	3	US-08-923-992A-8	Sequence 8, Appli
4	168	7.0	1164	3	US-08-923-992A-10	Sequence 10, Appl
5	167.5	7.0	957	4	US-09-914-259-16	Sequence 16, Appl
6	166	6.9	1104	3	US-08-923-992A-4	Sequence 4, Appli
7	166	6.9	1128	3	US-08-923-992A-6	Sequence 6, Appli
8	166	6.9	2662	4	US-09-595-684B-31	Sequence 31, Appl
9	163.5	6.8	956	4	US-09-914-259-17	Sequence 17, Appl
10	162	6.8	8991	4	US-08-714-741-32	Sequence 32, Appl
11	159.5	6.6	816	2	US-08-533-306A-6	Sequence 6, Appli
12	159.5	6.6	816	2	US-08-742-923A-6	Sequence 6, Appli
13	158.5	6.6	885	2	US-08-533-306A-4	Sequence 4, Appli
14	158.5	6.6	885	2	US-08-742-923A-4	Sequence 4, Appli
15	156.5	6.5	1087	4	US-09-914-259-12	Sequence 12, Appl
16	156	6.5	1354	3	US-08-685-871-2	Sequence 2, Appli
17	154.5	6.4	1231	4	US-08-714-741-41	Sequence 41, Appl
18	153.5	6.4	984	1	US-08-242-932-2	Sequence 2, Appli
19	153.5	6.4	984	1	US-08-714-481-2	Sequence 2, Appli
20	153.5	6.4	984	5	PCT-US95-06111-2	Sequence 2, Appli
21	153	6.4	1588	5	PCT-US93-07261-11	Sequence 11, Appl
22	153	6.4	1663	5	PCT-US93-07261-16	Sequence 16, Appl
23	153	6.4	2285	4	US-09-308-375-2	Sequence 2, Appli
24	150.5	6.3	1010	4	US-09-134-001C-5178	Sequence 5178, Ap
25	147	6.1	534	4	US-09-103-664A-2	Sequence 2, Appli
26	146.5	6.1	1388	2	US-08-685-576-1	Sequence 1, Appli
27	146.5	6.1	1388	2	US-08-685-576-4	Sequence 4, Appli
28	145.5	6.1	652	4	US-08-559-896B-2	Sequence 2, Appli
29	145	6.0	588	4	US-08-714-741-42	Sequence 42, Appl
30	144.5	6.0	343	3	US-08-937-271-17	Sequence 17, Appl
31	144	6.0	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
32	143.5	6.0	864	4	US-08-714-741-40	Sequence 40, Appl
33	143	6.0	3878	4	US-09-914-259-11	Sequence 11, Appl
34	142	5.9	967	4	US-09-914-259-21	Sequence 21, Appl
35	142	5.9	1032	4	US-09-914-259-26	Sequence 26, Appl
36	141.5	5.9	1146	4	US-08-914-999-6	Sequence 6, Appli
37	141	5.9	1507	3	US-08-929-329-5	Sequence 5, Appli
38	139.5	5.8	963	4	US-09-914-259-20	Sequence 20, Appl
39	139	5.8	1939	4	US-09-310-187A-1	Sequence 1, Appli
40	139	5.8	3111	2	US-08-460-309-4	Sequence 4, Appli
41	139	5.8	3111	2	US-08-125-077-4	Sequence 4, Appli
42	138	5.8	630	3	US-08-973-462-9	Sequence 9, Appli
43	138	5.8	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
44	138	5.8	1234	4	US-09-592-054-8	Sequence 8, Appli
45	137.5	5.7	586	2	US-08-630-822A-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8

Query Match 7.5%; Score 179; DB 3; Length 1786;
Best Local Similarity 21.5%; Pred. No. 0.00012;
Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;

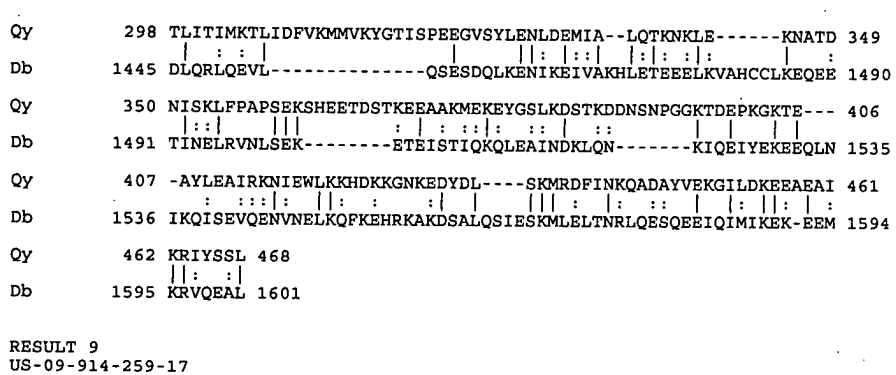
```
Qy          27 QDKSLHNRELSAE--RPLNEQIAEAEEDKIKKTYPPENKPGQSN-------YSFVDNL-- 75
               :: :: : |  ||    ||| :     ||:::|     :  ::|           ||:
Db         957 EENAVESNENVAENLEKLNETVFNTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKG 1016

Qy          76 ---NLLK----------AITEKEKIEKERQSIRSSPLDNKLN--------VEDVDSTK-- 112
                  |||            |  :||::   ::: || |||  |        :|:: ||:
Db        1017 IQENLLTGMFRSIETSIVIQSEEKVDL-NENVVSSILDNIENMKEGLLNKLENISSTEGV 1075

Qy         113 --------NRKLIDDYD-----------STKSGLDHKFQDDPDGLHQLDGTPLTAEDI- 151
                     : :  | |              :   ||   | :  | : : :   :| |:|
Db        1076 QETVTEHVEQNVYVDVDVPAMKDQFLGILNEAGGLKEMFFNLED-VFKSESDVITVEEIK 1134

Qy         152 -------VHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKE 204
                    | |    | || :  : |           : | :   | |:: : ::: |
Db        1135 DEPVQKEVEKETVSIIEEMEENIVD----------VLEEEKEDLTDKMIDAVEESIEIS 1183

Qy         205 ANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTK 264
               ::: ||        ||:    | :|   :: : : :   : ||:|   | | |   |  |
Db        1184 SDSKEE--------TES----IKDKEKDVSLVVEEVQDNDMDESVEKVLELKNMEEELMK 1231

Qy         265 TYSEDN------FEELQYFPNFYA-LLKSIDSEKEAK------------EKETLITIMK 304
```

```
Db       1232 DAVEINDITSKLIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIIDAKDDTLEKVIE 1291

Qy        305 ------TLIDFVKMMVKYGTISPE--EGVSYLENLDEMIALQTK--NKLEKNATDNISKL 354

Db       1292 EEHDITTTLDEV---VELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESEILEDYKEL 1348

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRK 414

Db       1349 KTIETDILEEKKEIEKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLEEVHE-LKE 1407

Qy        415 NIEWLKKHDK--KGNKEDYDLSKMRDFINKQADAY---VEKGILDK----------EEA 458

Db       1408 EVEHIISGDAHIKGLEED-DLEEVDDLKGSILDMLKGDMELGDMDKESLEDVTTKLGERV 1466

Qy        459 EAIKRIYSS 467

Db       1467 ESLKDVLSS 1475


RESULT 2
US-08-923-992A-2
; Sequence 2, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1164 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-2

  Query Match             7.2%;  Score 172.5;  DB 3;  Length 1164;
  Best Local Similarity   21.6%;  Pred. No. 0.0002;
  Matches  124;  Conservative   77;  Mismatches  189;  Indels  185;  Gaps   25;

Qy         36 LSAERPLNE--------QIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT-EKEK 86

Db        149 LELENQFNETNRLLHIKQHEEVEKDKKAK----QQKTLKQSDTKVDLSNIDKELNHQKSQ 204

Qy         87 IEK--ERQSIRSSPLDNKL-------------NVEDVDSTKNRKLIDDYDSTKSGLDHK 130

Db        205 VEKMAEQKGITNEDKDSMLKKIEDIRKQAQQADKKEDAEVKVREELGKLFSSTKAGLDQE 264

Qy        131 FQDDPDGLHQLDGTPLTAEDIVHKI-----------AARIYEENDRAVFDKIVSKLLNLG 179

Db        265 IQE-----HVKKET--SSEENTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQF 317

Qy        180 L-------------------ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEE 210

Db        318 LENAQKLKEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEG 377

Qy        211 DPNKPTSWTE-------------------NQAGKIPEKVTPMAAIQDGLAKGENDETVSN 251

Db        378 KLNQSKNLPELKQLEEEAHSKLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQ---Q 434

Qy        252 TLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVK 311

Db        435 KIELTVSPENITVYEGED--------VKFTVTAKS-DS-------------KTTLDFSD 471

Qy        312 MMVKYG-TISPEEGVSYLENLD-----------------EMIALQTK----NKLEKNATD 349

Db        472 LLTKYNPSVSDRISTNYKTNTDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTI 531

Qy        350 NISKLFPAPSEKSHEETDSTKEEAA-----------------------KMEKEYGSLKD 385

Db        532 TVQKKEEKQVPKTPEQKDSKTEEKVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELME 591

Qy        386 STKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKK----HDKKGNKEDYDLSKMRDFIN 441

Db        592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMN 648

Qy        442 KQADAYVEKGILDKEEAE---------AIKRIYSS 467

Db        649 YQLHAQME--MLTRKVVQYMNKYPDNAEIKKIFES 681


RESULT 3
US-08-923-992A-8
; Sequence 8, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  8:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1098 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-8

  Query Match             7.0%;  Score 169;  DB 3;  Length 1098;
  Best Local Similarity   19.3%;  Pred. No. 0.00034;
  Matches  104;  Conservative   80;  Mismatches  182;  Indels  174;  Gaps   19;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLLKAI 81
                 | ::| |   : : :  ::  |   |:    ||   |     :        : | |  |
Db         74 GKREKQLQQWKNNLKNDVDNTILSHEQKNEFKTKIDETNDSDALLELENQFNETNRLLHI 133

Qy         82 TEKEKIEKERQSIRSSPL---DNKLNVEDVDSTKNR----------------KLIDD-- 119
                : |::||:::: :   |   | |::: ::|   |                | |:|
Db        134 KQHEEVEKDKKAKQQKTLKQSDTKVDLSNIDKELNHQKSQEAGITNEDKDSMLKKIEDIR 193

Qy        120 ----------------------YDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKI- 155
                                   : |||:||| : |:     |    | ::|:   |:
Db        194 KQAQQPDKKEDAEVKVREELGKLFSSTKAGLDQEIQE-----HVKKET--SSEENTQKVD 246

Qy        156 ----------AARIYEENDRAVFDKIVSKLLNLGL------------------ITESQA 186
                       | :  || |:|  ::  ::: |  |                    : :: :
Db        247 EHYANSLQNLAQKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIKETNVKLYKAMS 306

Qy        187 HTLEDEVAEV-------LQKLISKE---ANNYEEDPNKPTSWTE---------------- 220
                :||    |:       |: |::|      ||   |:  :  |
Db        307 ESLEQVEKELKHNSEANLEDLVAKSKEIVREYEGKLNQSKNLPELKQLEEEAHSKLKQVV 366

Qy        221 ---NQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYF 277
                 :  |  |:|||   ::  ||  ||::     : ||   |  |    ||
Db        367 EDFRKKFKTSEQVTPKKRLKRDLAANENNQ---QKIELTVSPENITVYEGED-------- 415

Qy        278 PNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYG-TISPEEGVSYLENLD---- 332
                |   || ||                || :||  :: ||  ::|     :|  | |
Db        416 VKFTVTAKS-DS-------------KTTLDFSDLLTKYNPSVSDRISTNYKTNTDNHKI 460

Qy        333 -------------EMIALQTK----NKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK 375
                          : : |: |    | :||  |  : |       |: |: ||  ||
Db        461 AEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTITVQKKEEKQVPKTPEQKDSKTEEKVP 520

Qy        376 MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSK 435
                | :             ||   : |     :  || : | |: | :  :  :  :| : |
Db        521 QEPK------------SNDKNQLQELIKSAQQELEKLEKAIKELMEQPEIPSNPEYGIQK 568


RESULT 4
US-08-923-992A-10
; Sequence 10, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1164 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-10

  Query Match             7.0%;  Score 168;  DB 3;  Length 1164;
  Best Local Similarity   19.0%;  Pred. No. 0.00044;
  Matches  121;  Conservative   85;  Mismatches  177;  Indels  254;  Gaps   25;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLLKAI 81
                 | ::| |   : : :  ::  |   |:    ||   |     :        : | |  |
Db        105 GKREKQLQQWKNNLKNDVHNTILSHEQKNEFKTKIDETNDSDALLELENQFNETNRLLHI 164

Qy         82 TEKEKIEKERQSIRSSPL---DNKLNVEDVDSTKNR----------------------K 115
                : |::||:::: :   |   | |::: ::|   |                        |
Db        165 KQHEEVEKDKKAKQQKTLKQSDTKVDLSNIDKELNHQKSQVETMAEQLGITNEDKDSMLK 224

Qy        116 LIDD-------------------------YDSTKSGLDHKFQDDPDGLHQLDGTPLTAED 150
                |:|                       : |||:||| : |:     |    |  |:|:
Db        225 KIEDIRKQAQQADKKEDAEVKVREELGKLFTSTKAGLDQEIQE-----HVKKET--TSEE 277

Qy        151 IVHKI-----------AARIYEENDRAVFDKIVSKLLN------------LGLITE---- 183
                 |:           | :  || |:|  ::  ::: |                 || |
Db        278 NTQKVDEHYPNSLQNLAQKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIKETNVK 337

Qy        184 ------------------SQAHTLEDEVAEVLQKLISKE-ANNYEEDPNKPTSWTE--- 220
                             ||| ||| ||:      |||    ||   |:  :  |
Db        338 LYKAMSESLEQVEKQLKHNSQAN-LEDLVAK------SKEIVREYEGKLNQSKNLPELKQ 390

Qy        221 ----------------NQAGKIPEKVTPMAAIQDGLAKGENDE-----TVS-NTLTLTNG 258
                             :  |  |:|||   ::  ||  ||::     |||   :|:  |
Db        391 LEEEAHSKLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQQKIELTVSPENITVYEG 450

Qy        259 LERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYG- 317
                        |:|:     : |    ||                || :||  :: ||
Db        451 ------------EDLK-----FTLTAKSDS-------------KTTLDFSDLLTKYNP 478

Qy        318 TISPEEGVSYLENLD-----------------EMIALQTK----NKLEKNATDNISKLFP 356
                ::|     :|  | |               : :  |: |     | ::|   |  : |
Db        479 SVSDRISTNYKTNTDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVQKTFTITVQKKEE 538

Qy        357 APSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNI 416
                  |: |: ||  ||     | :             ||   :  |      :  || : | |
Db        539 KQVPKTPEQKDSKTEEKVPQEPK------------SNDKNQLQELIKSAQQQLEKLEKAI 586

Qy        417 EWLKKHDKKGNKEDYDLSK-----------------------------------MRDF 439
                : | :  :  :  :| : |                                   ||
Db        587 KELMEQPEIPSNPEYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDF 646

Qy        440 INKQADAYVEKGILDKEEAE---------AIKRIYSS 467
                :| |  | :|  :| ::  :          ||:|: |
Db        647 MNYQLHAQME--MLTRKVVQYINKYPDNAEIKKIFES 681


RESULT 5
US-09-914-259-16
; Sequence 16, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
```

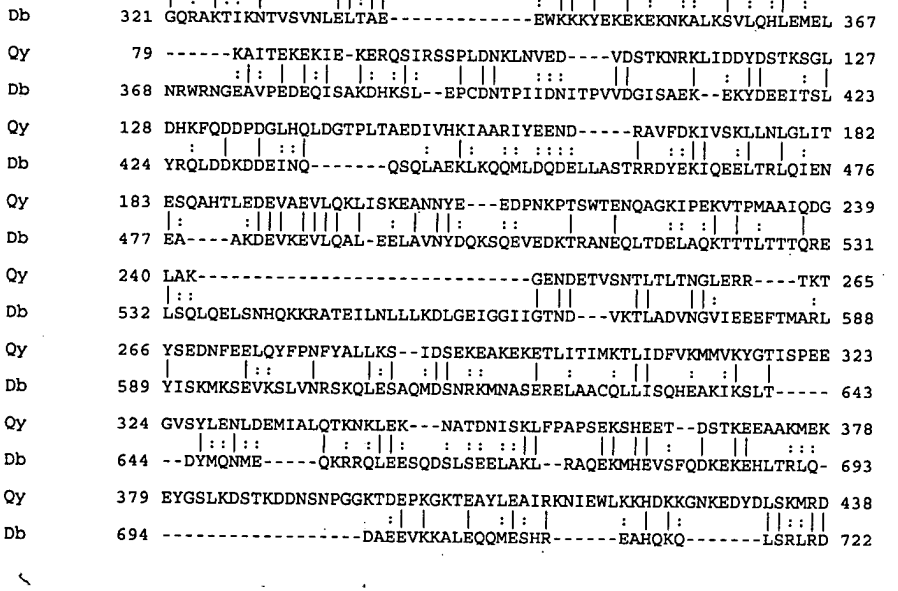
```
;  APPLICANT: Makowski, Lee
;  APPLICANT:  Hyman, Paul
;  APPLICANT:  Williams, Mark
;  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
;  FILE REFERENCE: 8471-010-999
;  CURRENT APPLICATION NUMBER: US/09/914,259
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 180
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 957
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-914-259-16

  Query Match             7.0%;  Score 167.5;  DB 4;  Length 957;
  Best Local Similarity   20.3%;  Pred. No. 0.00036;
  Matches  103;  Conservative   98;  Mismatches  182;  Indels  125;  Gaps   23;

Qy         25 GSQDKSLHNR-----ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLL- 78
Db        321 GQRAKTIKNTVSVNLELTAE------------EWKKKYEKEKEKNKTLKNVIQHLEMEL 367

Qy         79 ------KAITEKEKIEKERQSIRSSPLDNKLNVEDVD------STKNRKLIDDYDSTKSG 126
Db        368 NRWRNGEAVPEDEQISAKDQK-NLEPCDNTPIIDNIAPVVAGISTEEK---EKYDEEISS 423

Qy        127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEND-----RAVFDKIVSKLLNLGLI 181
Db        424 LYRQLDDKDDEINQ-------QSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIE 476

Qy        182 TESQAHTLEDEVAEVLQKLISKEANNYE---EDPNKPTSWTENQAGKIPEKVTPMAAIQD 238
Db        477 NEA----AKDEVKEVLQAL-EELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQR 531

Qy        239 GLAK-----------------------------GENDETVSNTLTLTNGLERR----TK 264
Db        532 ELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTND---VKTLADVNGVIEEEFTMAR 588

Qy        265 TYSEDNFEELQYFPNFYALLKS--IDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPE 322
Db        589 LYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLT---- 644

Qy        323 EGVSYLENLDEMIALQTKNKLEK---NATDNISKLFPAPSEKSHEETDSTKE-------- 371
Db        645 ---DYMQNME-----QKRRQLEESQDSLSEELAKL--RAQEKMHEVSFQDKEKEHLTRLQ 694

Qy        372 EAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDY 431
Db        695 DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEK-LSSDY 753

Qy        432 DLSKMRDFINKQADAYVEKGIL--DKEE 457
Db        754 NKLKIED---QEREMKLEKLLLLNDKRE 778


RESULT 6
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1104 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-4

  Query Match             6.9%;  Score 166;  DB 3;  Length 1104;
  Best Local Similarity   21.2%;  Pred. No. 0.00057;
  Matches  121;  Conservative   80;  Mismatches  190;  Indels  181;  Gaps   26;

Qy         17 IQAFPKPGGSQDKSLHNREL-----SAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF 71
Db        139 VEKHNKP--NQQKTLKQSDTKVDLSNIDKELNHQKSQVEA----------MAEQAGITN 185

Qy         72 VDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKF 131
Db        186 EDKDSMLKKI---EDIRKQAQQ-----ADKK---EDAEVKVREELGKLFSSTKAGLDQQI 234

Qy        132 QDDPDGLHQLDGTPLTAEDIVHKI----------AARIYEENDRAVFDKIVSKLLNLGL 180
Db        235 QE-----HVKKET--SSEENTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQFL 287

Qy        181 -------------------ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEED 211
Db        288 ENAQKLKEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLQDLVAKSKEIVREYEGK 347

Qy        212 PNKPTSWTE-------------------NQAGKIPEKVTPMAAIQDGLAKGENDETVSNT 252
Db        348 LNQSKNLPELKQLEEEAHSKLKQVVEHFRKKFKTSEQVTPKKRVKRDLAANENNQ---QK 404

Qy        253 LTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKM 312
Db        405 IELTVSPENITVYEGED--------VKFTVTAKS-DS-------------KTTLDFSDL 441

Qy        313 MVKYG-TISPEEGVSYLENLD-----------------EMIALQTK----NKLEKNATDN 350
Db        442 LTKYNPSVSDRISTNYKTNTDNHKIAEITIKNLKLNQSQTVTLKAKDDSGNVVEKTFTIT 501

Qy        351 I----SKLFPAPSEKSHEETD--------------------STKEEAAKMEKEYGSLKDS 386
Db        502 VQKKEEKQVPKTPEQKHSKTEQNVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELMEQ 561

Qy        387 TKDDNSNP--GGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQA 444
Db        562 -PEIPSNPEYGIQKSIWESQKEPIQEAITSFNKIIGDSSSKYYTEHYFNKYKSHFMNYQL 620

Qy        445 DAYVEKGILDKEEAE---------AIKRIYSS 467
Db        621 HAQME--ILTRKVVQYMNKYPDNAEIKKIFES 650


RESULT 7
```

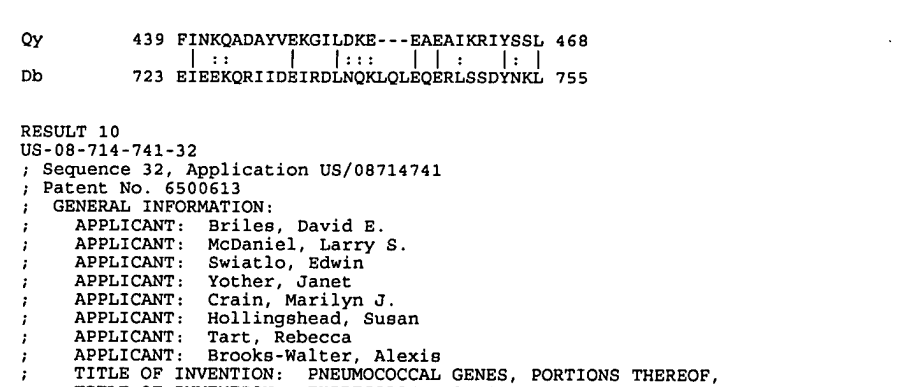
```
US-08-923-992A-6
; Sequence 6, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1128 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-6

  Query Match             6.9%;  Score 166;  DB 3;  Length 1128;
  Best Local Similarity   19.3%;  Pred. No. 0.00059;
  Matches  119;  Conservative   86;  Mismatches  200;  Indels  210;  Gaps   23;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLKAI 81
              | ::| |   : : :  ::  |   |:     ||   |      :        :  | |  |
Db         69 GKREKQLQQWKNNLKNDVDNTILSHEQKNEFKTKIDETNDSDALLELENQFNETNRLLHI 128

Qy         82 TEKEKIEKERQSIRSSPL---DNKLNVEDVDSTKNR----------------------K 115
              . : |::||:::: :    |    | |::: ::|    |                     |
Db        129 KQHEEVEKDKKAKQQKTLKQSDTKVDLSNIDKELNHQKSPVEKMAEPKGITNEDKDSMLK 188

Qy        116 LIDD-------------------------YDSTKSGLDHKFQDDPDGLHQLDGTPLTAED 150
               |:|                        : |||:||| :       :|:      ::|:
Db        189 KIEDIRKQAQQADKKEDAEVKVREELGKLFSSTKAGLDQE-------IHEHVKKETSSEE 241

Qy        151 IVHKI-----------AARIYEENDRAVFDKIVSKLLNLGL------------------ 180
                 |:           | :  || |:|  ::  ::: |  |
Db        242 NTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQFLENAQKLKEMQPLIKETNVK 301

Qy        181 ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEEDPNKPTSWTE---------- 220
              : :: : :||    |:       |: |::|       ||   |:  :  |
Db        302 LYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEGKLNQSKNLPELKQLEEEAHS 361

Qy        221 ---------NQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNF 271
                      :  |  |:|||   ::  ||  ||::     : ||   |  |    ||
Db        362 KLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQ---QKIELTVSPENITVYEGED-- 416

Qy        272 EELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYG-TISPEEGVSYLEN 330
                    |    || ||               || :||  :: ||  ::|    :|  |

Db        417 ------VKFTVTAKS-DS--------------KTTLDFSDLLTKYNPSVSDRISTNYKTN 455

Qy        331 LD----------------EMIALQTK----NKLEKNATDNISKLFPAPSEKSHEETDST 369
                |                : : |: |    | :||  |  : |        |: |: ||
Db        456 TDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTITVQKKEEKQVPKTPEQKDSK 515

Qy        370 KEEAA-----------------------KMEKEYGSLKDSTKDDNSNPGGKTDEPKGKT 405
                ||                      |:||    | :   :  |||  :    |
Db        516 TEEKVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELMEQ-PEIPSNP--EYGIQKSIW 572

Qy        406 EAYLEAIRKNIEWLKK----HDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE-- 459
              |:  | |:: |   ||        |   | |       ||:| |  | :|  :| ::  :
Db        573 ESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMNYQLHAQME--MLTRKVVQYM 630

Qy        460 -------AIKRIYSS 467
                      ||:|: |
Db        631 NKYPDNAEIKKIFES 645


RESULT 8
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Ohashi, Cara
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  VAisberg, Eugeni
;  APPLICANT:  Wood, Kenneth
;  APPLICANT:  Yu, Ming
;  TITLE OF INVENTION: Human kinesins and methods of producing
;  TITLE OF INVENTION:  and purifying human kinesins
;  FILE REFERENCE: cytop036
;  CURRENT APPLICATION NUMBER: US/09/595,684B
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 09/295,612
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 2662
;   TYPE: PRT
;   ORGANISM: Human
US-09-595-684B-31

  Query Match             6.9%;  Score 166;  DB 4;  Length 2662;
  Best Local Similarity   21.8%;  Pred. No. 0.0019;
  Matches  119;  Conservative  103;  Mismatches  175;  Indels  150;  Gaps   28;

Qy         26 SQDKSLHNRELSAERPLN---EQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT 82
              :|:|   |  :  |  |:   :::|| ||   :|:   : |  |         ||: : ::
Db       1101 AQEK---NHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQL-------LNVQEEMS 1150

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTK---NRKLIDDYDSTKS---------GLDHK 130
              | :|   | :::::   : :| :| ::: :    :|| ::|:  ||           |
Db       1151 EMQKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVLKELQKS 1210

Qy        131 FQDDPDGL----HQLDGTPL-TAEDI----VH-KIAARIYEENDRAVFDKIVSKLLNLGL 180
              |: : | |      ::: | | | |::     :| |       :|  |:| :|  ::::|
Db       1211 FETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEK-TAQIINTQD 1269

Qy        181 ITESQAHTLEDEVAEVL---QKLIS--KEANNYEEDPNKPTSWTENQAGKIPEKVTPMAA 235
              : :|  ||   |   ||   |:|:   |: :  :|  |:    ||    |     | :|
Db       1270 LEKS--HTKLQEEIPVLHEEQELLPNVKKVSETQETMNELELLTEQSTTK---DSTTLAR 1324

Qy        236 IQ----------------DGLAK-GENDETVSNTL-------------TLTNGLERRTK 264
              |:                | |  :| :|:   |                ||    | ::|
Db       1325 IEMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSK 1384

Qy        265 TYSEDNFEEL----------QYFPNFYALLK-------------SIDSEKE-AKEKE 297
                  | :|            |: |   |||:              | |  |  ||||:
Db       1385 QEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKD 1444
```



RESULT 9
US-09-914-259-17
; Sequence 17, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-914-259-17

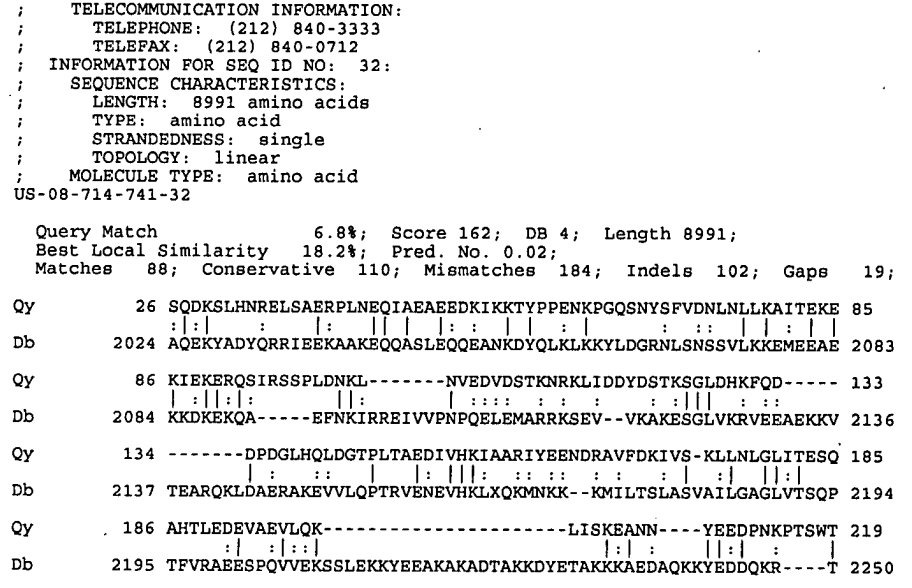
Query Match 6.8%; Score 163.5; DB 4; Length 956;
Best Local Similarity 20.7%; Pred. No. 0.00072;
Matches 106; Conservative 91; Mismatches 169; Indels 147; Gaps 25;





RESULT 10
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; TITLE OF INVENTION: PORTIONS AND PRODUCTS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8991 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-714-741-32

Query Match 6.8%; Score 162; DB 4; Length 8991;
Best Local Similarity 18.2%; Pred. No. 0.02;
Matches 88; Conservative 110; Mismatches 184; Indels 102; Gaps 19;



```
Qy        220 ENQAGKIPE---KVTPMA-----------AIQDGLAKGENDETVSNTLT-LTNGLERRTK 264
Db       2251 EEKARKEAEASQKLIDVALVVQNAYKEYREVQNQRSKYKSDADYQKKLTEVDSKIEKARK 2310

Qy        265 TYS--EDNFEELQYFPNFYALL----KSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGT 318
Db       2311 EQQDLQNNFNEVR------AVVAPDPTCVGXDXRMNKKKMILTSLASV-----AILGAGX 2359

Qy        319 ISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAKMEK 378
Db       2360 VTSQPTXVRAEEAPQVV---EKSSLEKKYEEAKAKY-----DAAKKDYDEAKKKAAEAQK 2411

Qy        379 EYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGN---KEDYDLSK 435
Db       2412 KYEEDQKKTEEKAEKAKAASEEIAKATEEVQKAVLDYITAIRNHNDSGKTSAEEAENKAK 2471

Qy        436 MRDF 439
Db       2472 ERDY 2475
```

RESULT 11
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 6.6%; Score 159.5; DB 2; Length 816;
Best Local Similarity 22.6%; Pred. No. 0.0011;
Matches 109; Conservative 81; Mismatches 193; Indels 99; Gaps 21;

```
Qy         26 SQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITE 83
Db        158 SHREEMENEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTK 215

Qy         84 KEKIEKERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KF 131
Db        216 LRQLEEERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRF 275

Qy        132 QDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQA 186
Db        276 QKEIENLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQ 328

Qy        187 HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGEN 245
Db        329 RKFDQLLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEE 379

Qy        246 DETVSNTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLIT 301
Db        380 LERTNKMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQME 420

Qy        302 IMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------L 354
Db        421 EMKTQLEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQL 476

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAI 412
Db        477 HEYETELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDF 535

Qy        413 RKNIE-----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDK 455
Db        536 QRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEK 588

Qy        456 EE 457
Db        589 EE 590
```

RESULT 12
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: No. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```
;    MOLECULE TYPE:  protein
US-08-742-923A-6

  Query Match             6.6%;  Score 159.5;  DB 2;  Length 816;
  Best Local Similarity   22.6%;  Pred. No. 0.0011;
  Matches  109;  Conservative   81;  Mismatches  193;  Indels   99;  Gaps   21;

Qy         26 SQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITE 83
              |  : : |   |    |||  || :  |: |      :   |      : :    |   |:
Db        158 SHREEMENEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTK 215

Qy         84 KEKIEKERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KF 131
               ::|:|| |::      :: | |:|   || |      :| : |: ||   |:    :|
Db        216 LRQLEEERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRF 275

Qy        132 QDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQA 186
              | : : | |       | | : |   |: :| |  | |     ::|| |       | :
Db        276 QKEIENLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQ 328

Qy        187 HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGEN 245
                :  :||  :| ||   : | :: ::     | :| :   |   :| |::: |   |
Db        329 RKFDQLLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEE 379

Qy        246 DETVSNTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLIT 301
              |  :  |     |      |:|    | ||              ||  :  || :
Db        380 LERTNKMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQME 420

Qy        302 IMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------L 354
              |||  :: ::  ::    : |:    ||   : :  | :  |:      |  |       |
Db        421 EMKTQLEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQL 476

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAI 412
                :|   |  :     ||| : | | |||     :|   |: :  |   | :| ::
Db        477 HEYETELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDF 535

Qy        413 RKNIE-----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDK 455
              :: :|           |:::||    : || :::: :       ||||       |:|
Db        536 QRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEK 588

Qy        456 EE 457
              ||
Db        589 EE 590


RESULT 13
US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869COB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-533-306A-4

  Query Match             6.6%;  Score 158.5;  DB 2;  Length 885;
  Best Local Similarity   22.9%;  Pred. No. 0.0015;
  Matches  109;  Conservative   79;  Mismatches  190;  Indels   99;  Gaps   21;

Qy         31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITEKEKIE 88
              | |   |    |||  || :  |: |      :   |      : :    |   |: ::|
Db        232 LQNEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTKLRQLE 289

Qy         89 KERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KFQDDPD 136
              :|| |::      :: | |:|   || |      :| : |: ||   |:    :|| : :
Db        290 EERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRFQKEIE 349

Qy        137 GLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQAHTLED 191
              | |       | | : |   |: :| | | | |     ::|| |       | :    :
Db        350 NLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQRKFDQ 402

Qy        192 EVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGENDETVS 250
              :|| :| ||   : | :: ::     | :| :   |   :| |::: |   | |  :
Db        403 LLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEELERTN 453

Qy        251 NTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTL 306
                |       |      |:|    | ||               ||  :  || :  |||
Db        454 KMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQMEEMKTQ 494

Qy        307 IDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------LFPAPS 359
              :: ::  ::    : |:    ||   : :  | :  |:      |  |       |    :
Db        495 LEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYET 550

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAIRKNIE 417
              |   |  :     ||| : | | |||     :|   |: :  |   | :| ::  :: :|
Db        551 ELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELE 609

Qy        418 -----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDKEE 457
                        |:::||    : || :::: :       ||||       |:|||
Db        610 DARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEKEE 659


RESULT 14
US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,923A
;      FILING DATE:  No. 5869611ember 1, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869DVC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-742-923A-4

  Query Match             6.6%;  Score 158.5;  DB 2;  Length 885;
  Best Local Similarity   22.9%;  Pred. No. 0.0015;
  Matches  109;  Conservative   79;  Mismatches  190;  Indels   99;  Gaps   21;

Qy         31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITEKEKIE 88
Db        232 LQNEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTKLRQLE 289

Qy         89 KERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KFQDDPD 136
Db        290 EERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRFQKEIE 349

Qy        137 GLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQAHTLED 191
Db        350 NLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQRKFDQ 402

Qy        192 EVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGENDETVS 250
Db        403 LLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEELERTN 453

Qy        251 NTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTL 306
Db        454 KMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQMEEMKTQ 494

Qy        307 IDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------LFPAPS 359
Db        495 LEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYET 550

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAIRKNIE 417
Db        551 ELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELE 609

Qy        418 -----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDKEE 457
Db        610 DARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEKEE 659


RESULT 15
US-09-914-259-12
; Sequence 12, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
;  APPLICANT: Makowski, Lee
;  APPLICANT:  Hyman, Paul
;  APPLICANT:  Williams, Mark
;  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
;  FILE REFERENCE: 8471-010-999
;  CURRENT APPLICATION NUMBER: US/09/914,259
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 180
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;  LENGTH: 1087
;  TYPE: PRT
;  ORGANISM: Oryctolagus cuniculus
US-09-914-259-12

  Query Match             6.5%;  Score 156.5;  DB 4;  Length 1087;
  Best Local Similarity   19.3%;  Pred. No. 0.0028;
  Matches  103;  Conservative   83;  Mismatches  180;  Indels  167;  Gaps   22;

Qy         31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLLKAITEKEKIEKE 90
Db        304 IHEKEVEIDR-LNEQIIKLQQQ-----------------------LKITTDNKVIEEK 337

Qy         91 RQSIRSSPLDNKLNVEDVDST-KNR--------KLIDDYDSTKSGLDHKFQDDPDGLHQL 141
Db        338 NELIRDLEAQIECLMSDQERVRKNREEEIEQLNEVIEKLQQELANIDQKTSVDPSSLSE- 396

Qy        142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVA------- 194
Db        397 -----EADSLKHQL-------------DKVIAEKLAL----EHQVETTNEEMAVTKNVLK 434

Qy        195 -------EVLQKLIS--KEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAI--QDGLAKG 243
Db        435 ETNFKMNQLTQELCSLKRERSKMERIQSVPEKSVNMSVGDLSKDKPEMDLIPTEDALAQL 494

Qy        244 E-------NDETVSNTL-TLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKE 295
Db        495 ETQTQLRSSEESSKVSLSSLETKLLQLESTVSTKDLELTQCY-------KQIQDMREQGR 547

Qy        296 KET------LITIMKTL-----------------IDFVKMMVKYGTISPEEG-------- 324
Db        548 SETEMLQTKIVSLQKVLEEKVAAALVSQVQLEAVQEYVKLCADKPAVSSDPARTEVPGLS 607

Qy        325 -----------------VSYLENL-----------DEMIALQTKNKLEK-NATDNISKLF 355
Db        608 QLAGNTMESDVSALTWRISELESQLVEMHSSLISEKEQVEIAEKNALEKEKKLQELQKLV 667

Qy        356 PAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPG------GKTDEPKGKTEAYL 409
Db        668 QDSETKQRE-----RERQSRLHGDLGVLESTTSEESGVFGELEALRAESAAPKGELANYK 722

Qy        410 EAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADA---YVEKGILDKEEAE 459
Db        723 ELAEKLQEELLVKETNMASLPKELSHVRDQLTEAEDKLSHFSEKE--DKTEVQ 773


Search completed: January  2, 2004, 12:36:20
Job time : 22 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:34:26 ; Search time 39 Seconds
(without alignments)
2407.695 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2399	100.0	468	11	US-09-931-836-67	Sequence 67, Appl
2	2399	100.0	468	12	US-10-035-977-67	Sequence 67, Appl
3	2399	100.0	468	12	US-10-063-735-150	Sequence 150, App
4	2399	100.0	468	12	US-10-137-870-90	Sequence 90, Appl
5	2399	100.0	468	12	US-10-140-018-90	Sequence 90, Appl
6	2399	100.0	468	12	US-10-140-021-90	Sequence 90, Appl
7	2399	100.0	468	12	US-10-140-274-90	Sequence 90, Appl
8	2399	100.0	468	12	US-10-140-471-90	Sequence 90, Appl
9	2399	100.0	468	12	US-10-140-807-90	Sequence 90, Appl
10	2399	100.0	468	12	US-10-140-922-90	Sequence 90, Appl
11	2399	100.0	468	12	US-10-140-924-90	Sequence 90, Appl
12	2399	100.0	468	12	US-10-140-926-90	Sequence 90, Appl
13	2399	100.0	468	12	US-10-141-698-90	Sequence 90, Appl
14	2399	100.0	468	12	US-10-141-702-90	Sequence 90, Appl
15	2399	100.0	468	12	US-10-141-704-90	Sequence 90, Appl
16	2399	100.0	468	12	US-10-142-421-90	Sequence 90, Appl
17	2399	100.0	468	12	US-10-142-432-90	Sequence 90, Appl
18	2399	100.0	468	12	US-10-142-767-90	Sequence 90, Appl
19	2399	100.0	468	12	US-10-143-033-90	Sequence 90, Appl
20	2399	100.0	468	12	US-10-144-994-90	Sequence 90, Appl
21	2399	100.0	468	12	US-10-145-628-90	Sequence 90, Appl
22	2399	100.0	468	12	US-10-145-631-90	Sequence 90, Appl
23	2399	100.0	468	12	US-10-145-633-90	Sequence 90, Appl
24	2399	100.0	468	12	US-10-145-746-90	Sequence 90, Appl
25	2399	100.0	468	12	US-10-145-748-90	Sequence 90, Appl
26	2399	100.0	468	12	US-10-145-823-90	Sequence 90, Appl
27	2399	100.0	468	12	US-10-145-826-90	Sequence 90, Appl
28	2399	100.0	468	12	US-10-145-870-90	Sequence 90, Appl
29	2399	100.0	468	12	US-10-145-876-90	Sequence 90, Appl
30	2399	100.0	468	12	US-10-145-959-90	Sequence 90, Appl
31	2399	100.0	468	12	US-10-146-724-90	Sequence 90, Appl
32	2399	100.0	468	12	US-10-146-725-90	Sequence 90, Appl
33	2399	100.0	468	12	US-10-146-795-90	Sequence 90, Appl
34	2399	100.0	468	12	US-10-147-495-90	Sequence 90, Appl
35	2399	100.0	468	12	US-10-147-501-90	Sequence 90, Appl
36	2399	100.0	468	12	US-10-147-504-90	Sequence 90, Appl
37	2399	100.0	468	12	US-10-147-506-90	Sequence 90, Appl
38	2399	100.0	468	12	US-10-147-509-90	Sequence 90, Appl
39	2399	100.0	468	12	US-10-147-510-90	Sequence 90, Appl
40	2399	100.0	468	12	US-10-147-511-90	Sequence 90, Appl
41	2399	100.0	468	12	US-10-147-529-90	Sequence 90, Appl
42	2399	100.0	468	12	US-10-152-397-90	Sequence 90, Appl
43	2399	100.0	468	12	US-10-153-586-90	Sequence 90, Appl
44	2399	100.0	468	12	US-10-158-783-90	Sequence 90, Appl
45	2399	100.0	468	12	US-10-158-786-90	Sequence 90, Appl

ALIGNMENTS

RESULT 1
US-09-931-836-67
; Sequence 67, Application US/09931836
; Publication No. US20030027249A1
; GENERAL INFORMATION:
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3030R1C1
; CURRENT APPLICATION NUMBER: US/09/931,836
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/112514
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114140
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115552
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22

; PRIOR APPLICATION NUMBER: 60/125774
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125778
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125826
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/127035
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/127706
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/130359
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131272
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/132371
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132379
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132383
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/135750
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/138166
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/146970
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/162506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/644848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 09/747259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/816744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/854208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/854280
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 67
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-931-836-67

Query Match 100.0%; Score 2399; DB 11; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 2
US-10-035-977-67
; Sequence 67, Application US/10035977
; Publication No. US20030134327A1
; GENERAL INFORMATION:
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.

; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3030R1C10
; CURRENT APPLICATION NUMBER: US/10/035,977
; CURRENT FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/112514
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114140
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115552
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/125774
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125778
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125826
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/127035
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/127706
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/130359
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131272
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/132371
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132379
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132383
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/135750
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/138166
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/146970
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/162506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/644848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 09/747259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/816744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/854208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/854280
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 67
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-035-977-67

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
```

```
Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 3
US-10-063-735-150
; Sequence 150, Application US/10063735
; Publication No. US20030138882A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,735
;  CURRENT FILING DATE:  2002-05-08
;      Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 150
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-735-150

  Query Match           100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 4
US-10-137-870-90
; Sequence 90, Application US/10137870
; Publication No. US20030138883A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C155
;  CURRENT APPLICATION NUMBER: US/10/137,870
;  CURRENT FILING DATE:  2002-05-03
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-137-870-90

  Query Match           100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

```
RESULT 5
US-10-140-018-90
; Sequence 90, Application US/10140018
; Publication No. US20030138885A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C158
;  CURRENT APPLICATION NUMBER: US/10/140,018
;  CURRENT FILING DATE:  2002-05-06
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-018-90

  Query Match           100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468

RESULT 6
US-10-140-021-90
; Sequence 90, Application US/10140021
; Publication No. US20030138886A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C167
;  CURRENT APPLICATION NUMBER: US/10/140,021
;  CURRENT FILING DATE:  2002-05-06
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-021-90

  Query Match           100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468

RESULT 7
US-10-140-274-90
```

; Sequence 90, Application US/10140274
; Publication No. US20030143674A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C161
; CURRENT APPLICATION NUMBER: US/10/140,274
; CURRENT FILING DATE: 2002-05-06
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-274-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 8
US-10-140-471-90
; Sequence 90, Application US/10140471
; Publication No. US20030138887A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C163
; CURRENT APPLICATION NUMBER: US/10/140,471
; CURRENT FILING DATE: 2002-05-06
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 9
US-10-140-807-90
; Sequence 90, Application US/10140807
; Publication No. US20030134354A1

```
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C174
;  CURRENT APPLICATION NUMBER: US/10/140,807
;  CURRENT FILING DATE:  2002-05-07
;  Prior Apploication removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-807-90

  Query Match            100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 10
US-10-140-922-90
; Sequence 90, Application US/10140922
; Publication No. US20030138889A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C179
;  CURRENT APPLICATION NUMBER: US/10/140,922
;  CURRENT FILING DATE:  2002-05-07
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-922-90

  Query Match            100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 11
US-10-140-924-90
; Sequence 90, Application US/10140924
; Publication No. US20030134355A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
```

```
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C177
;  CURRENT APPLICATION NUMBER: US/10/140,924
;  CURRENT FILING DATE:  2002-05-07
;  Prior Apploication removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-924-90

  Query Match             100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 12
US-10-140-926-90
; Sequence 90, Application US/10140926
; Publication No. US20030134356A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C187
;  CURRENT APPLICATION NUMBER: US/10/140,926
;  CURRENT FILING DATE:  2002-05-07
;  Prior Apploication removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-926-90

  Query Match             100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 13
US-10-141-698-90
; Sequence 90, Application US/10141698
; Publication No. US20030134357A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
```

```
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C206
;  CURRENT APPLICATION NUMBER: US/10/141,698
;  CURRENT FILING DATE:  2002-05-08
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-141-698-90

  Query Match            100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 14
US-10-141-702-90
; Sequence 90, Application US/10141702
; Publication No. US20030134358A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C208
;  CURRENT APPLICATION NUMBER: US/10/141,702
;  CURRENT FILING DATE:  2002-05-08
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-141-702-90

  Query Match            100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 15
US-10-141-704-90
; Sequence 90, Application US/10141704
; Publication No. US20030134359A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
```

; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C209
; CURRENT APPLICATION NUMBER: US/10/141,704
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-704-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

Search completed: January 2, 2004, 12:37:45
Job time : 40 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 20 Seconds
(without alignments)
2250.347 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1957	81.6	533	2	A37180	chromogranin/secre
2	196	8.2	2139	2	T18296	myosin heavy chain
3	191	8.0	2464	1	QRMSP1	microtubule-associ
4	186.5	7.8	2116	2	A26655	myosin heavy chain
5	184	7.7	665	2	B71609	hypothetical prote
6	183.5	7.6	1790	2	S67593	transport protein
7	180.5	7.5	2364	2	A56577	microtubule-associ
8	179.5	7.5	1526	2	A45605	mature-parasite-in
9	179	7.5	853	2	T51505	hypothetical prote
10	179	7.5	1005	2	A64465	hypothetical prote
11	179	7.5	1558	2	B71603	RESA-H3 antigen PF
12	179	7.5	2269	2	T28677	rhoptry protein -
13	178.5	7.4	1392	2	A43336	microtubule-vesicl
14	178.5	7.4	1427	2	S22695	restin - human
15	177	7.4	1871	2	D96796	probable heat shoc
16	172.5	7.2	1134	2	A60234	IgA Fc receptor pr
17	172.5	7.2	1164	1	FCSOAG	IgA Fc receptor pr
18	172.5	7.2	1979	2	C71622	hypothetical prote
19	171.5	7.1	3488	2	T34418	hypothetical prote
20	170	7.1	1804	2	T34518	nestin - golden ha
21	169	7.0	1272	2	C90593	hypothetical prote
22	168.5	7.0	746	2	T47237	myosin II heavy ch
23	168.5	7.0	926	2	T24923	hypothetical prote
24	168	7.0	2401	2	T28676	rhoptry protein -
25	167	7.0	976	2	T01553	hypothetical prote
26	166.5	6.9	1875	2	S38173	myosin-like protei
27	166.5	6.9	5327	2	T13564	microtubule-associ
28	166	6.9	991	2	H86168	hypothetical prote
29	166	6.9	1192	2	A71623	probable secreted
30	166	6.9	2663	1	S28261	centromere protein
31	164.5	6.9	719	2	A81358	hypothetical prote
32	164.5	6.9	911	2	S51441	hypothetical prote
33	164.5	6.9	1127	2	T28317	ORF MSV156 hypothe
34	164.5	6.9	1269	2	F84730	probable myosin he
35	164	6.8	1909	2	A45592	liver stage antige
36	163	6.8	1252	2	B42771	reticulocyte-bindi
37	163	6.8	2510	2	T28160	hypothetical prote
38	162.5	6.8	1738	2	T14867	interaptin - slime
39	162.5	6.8	2385	2	A32491	myosin heavy chain
40	162.5	6.8	2411	2	B32491	myosin heavy chain
41	161.5	6.7	1302	1	JC6009	surface-located me
42	161.5	6.7	1992	1	S02771	myosin heavy chain
43	161	6.7	852	2	D72230	conserved hypothet
44	160	6.7	1354	2	S74244	serine/threonine-s
45	159.5	6.6	821	2	S67087	hypothetical prote

ALIGNMENTS

RESULT 1
A37180
chromogranin/secretogranin-like vesicle protein precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 30-Sep-1993
C;Accession: A37180
R;Ottiger, H.P.; Battenberg, E.F.; Tsou, A.P.; Bloom, F.E.; Sutcliffe, J.G.
J. Neurosci. 10, 3135-3147, 1990
A;Title: 1B1075: a brain- and pituitary-specific mRNA that encodes a novel chromogranin/s
A;Reference number: A37180; MUID:90376160; PMID:2204688
A;Accession: A37180
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-533 <OTT>

Query Match 81.6%; Score 1957; DB 2; Length 533;
Best Local Similarity 86.7%; Pred. No. 3.5e-87;
Matches 383; Conservative 21; Mismatches 38; Indels 0; Gaps 0;

```
Qy         16 PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL 75
              |  :|||| |||||||||||||||||||||||||| |||||||| |:|| : |:| ||||
Db         63 PNSSFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKTYPSESKPSERNFSSVDNL 122

Qy         76 NLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDP 135
              |||||||||| :|| :||||||| ||:|||:| |||||||| |:|||||||| | ||||
Db        123 NLLKAITEKETVEKAKQSIRSSPFDNRLNVDDADSTKNRKLTDEYDSTKSGLDRKVQDDP 182

Qy        136 DGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAE 195
              |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db        183 DGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLNLGLITESQAHTLEDEVAE 242

Qy        196 VLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTL 255
              |||||||||||||| | |||| |||| |||||||||:|| |||    |||:||||||||
Db        243 ALQKLISKEANNYEEAPEKPTSRTENQDGKIPEKVTPVAATQDGFTNRENDDTVSNTLTL 302

Qy        256 TNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVK 315
              :|||||||  : :|:|||||||||||||| ||||||||||||||||||||||||||||||
Db        303 SNGLERRTNPHRDDDFEELQYFPNFYALLTSIDSEKEAKEKETLITIMKTLIDFVKMMVK 362

Qy        316 YGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK 375
              |||||||||||||||||| |||||||||||| ||: ||||||| ||||||||||||||||
Db        363 YGTISPEEGVSYLENLDETIALQTKNKLEKNTTDSKSKLFPAPPEKSHEETDSTKEEAAK 422

Qy        376 MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSK 435
              |||||||||||||||||| |||||| |||||||||||||||||||:|||||||||||||
Db        423 MEKEYGSLKDSTKDDNSNLGGKTDEAKGKTEAYLEAIRKNIEWLKKHNKKGNKEDYDLSK 482

Qy        436 MRDFINKQADAYVEKGILDKEE 457
              ||||||:|||||||||| :::
Db        483 MRDFINQQADAYVEKGILIRKK 504
```

RESULT 2
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C;Genetics:
A;Gene: mhcA
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 8.2%; Score 196; DB 2; Length 2139;
Best Local Similarity 21.0%; Pred. No. 0.1;
Matches 107; Conservative 108; Mismatches 191; Indels 104; Gaps 23;

```
Qy         20 FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPEN-------KPGQSNYSFV 72
Db        859 FEEEGKKKDKEIED--------LKKKLAE----EIKKREAAENALASATAKTGELEAKIQ 906

Qy         73 D----------------------NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVD- 109
Db        907 DLEDKISELESKLSAAELDKQELNLKIENLEEDKEELKETIDNLKGDLKDSKLKGEDLEV 966

Qy        110 ---------STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIY 160
Db        967 EITELNSQINTLN-ATVDKKDKTIAEMQESIDEKEDEITKLKGDIKLLEEEKDDL----- 1020

Qy        161 EENDR----AVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPT 216
Db       1021 -EQDRADVSATKDDIAKKLNKITI----ECEDAKDEIAKLEQELEDEE--NKNKDLTNEL 1073

Qy        217 SWTENQAGKIPEKVTPMAAI-------QDGLAKG-ENDETVSNTLTLTNG-LERRTKTYS 267
Db       1074 QQTQLKLGETEKSLAAQVAATKKASDERDTLSQNLENEKLTTKNLTKTKADLEKKISGLK 1133

Qy        268 EDNFEELQYFPNFYALLKSID-----SEKEAKEKETLITIMKTLIDFV-KMMVKYGTISP 321
Db       1134 QD-YEDLEDDKN------KIEGDLRNAQRKIKELDDEITKGADVSQYLQKQKEEY----- 1181

Qy        322 EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK----ME 377
Db       1182 ESQIAKMQEEKEAIGNDVKNK-EKTIKE--KELEIQSLQEKLDETEVEKEDAEKKKKEIE 1238

Qy        378 KEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL-SKM 436
Db       1239 KEMKALQEE-KENVESSKNSTEKDKKKLEDNLKDTQKKLDDMTADNEKLKAKAKDLEAQL 1297

Qy        437 RDFINKQADAYVEKGILDKEEAEAIKRIYS 466
Db       1298 NEVQDNHEKAVADAELLNKKKAQSDKELNS 1327
```

RESULT 3
QRMSP1
microtubule-associated protein MAP1B - mouse
N;Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
C;Accession: S07549; S44387; A33645
R;Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A;Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A;Reference number: A33645; MUID:90094539; PMID:2480963
A;Accession: S07549
A;Molecule type: mRNA
A;Residues: 1-2464 <NOB>
A;Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000
R;Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A;Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A;Reference number: S44387; MUID:94234720; PMID:8179328
A;Accession: S44387
A;Status: preliminary
A;Molecule type: protein
A;Residues: 653-663,'IC' <SAN>
C;Superfamily: microtubule-associated protein MAP1B
C;Keywords: microtubule binding; phosphoprotein; tandem repeat
F;589-786/Domain: microtubule binding #status experimental <MTB>
F;589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-694
R-K-E/D-X)
F;1861-2064/Region: 17-residue repeats
F;91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: phc
F;147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (cov
F;1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 8.0%; Score 191; DB 1; Length 2464;
Best Local Similarity 21.3%; Pred. No. 0.21;
Matches 118; Conservative 85; Mismatches 186; Indels 166; Gaps 24;

```
Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEA------EEDKIKKTYP--PENKPGQSNYSFVD 73
Db        550 KPVASKSVRKESKEETPEVTKTSQVEKTPKVESKEKVLVKKDKPVKTESKP--------- 600

Qy         74 NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQD 133
Db        601 ------SVTEKEVSSKEEQSPVKAEVAEKQATESKPKVTKDKVVKKEIKTK--LEEKKEE 652

Qy        134 DP--DGLHQLDGTPL------TAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQ 185
Db        653 KPKKEVVKKEDKTPLKKDEKPRKEEVKKEIKKEIKKE-ERKELKKEVKK--------ETP 703

Qy        186 AHTLEDEVAEVLQKLISKEANNYEEDPNK-------------PTSWTENQAGKIPE---- 228
Db        704 LKDAKKEVKKEEKKEVKKE----EKEPKKEIKKISKDIKKSTPQSDTKKPSALKPKVAKK 759

Qy        229 ----KVTPMAA--IQD-GLAK-----GENDETVSNTL---------------------- 253
Db        760 EESTKKEPLAAGKLKDKGKVKVIKKEGKTTEAAATAVGTAATTAAVVAAAGIAASGPVKE 819

Qy        254 -----TLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KET 298
Db        820 LEAERSLMSSPEDLTKDFEELKAEEIDVAKDIKPQLELIEDEEKLKETQPGEAYVIQKET 879

Qy        299 LITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL- 354
Db        880 EVS-------------KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFE 924

Qy        355 -----FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYL 409
Db        925 DEGAGFEESSETGDYEEKAETEEAEEPEED-GEDNASGSASKHSPTEDDESAKAEADVHL 983

Qy        410 EAIRKNI------------EWLKK-----HDKKGNKEDY-----------DLSKMRDFIN 441
Db        984 KEKRESVVSGDDRAEEDMDDVLEKGEAEQSEEEGEEEDKAEDAREEGYEPDKTEAEDYVM 1043

Qy        442 KQADAYVEKGILDKE 456
Db       1044 AVADKAAEAGVTEEQ 1058
```

RESULT 4
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Dictyostelium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
A;Reference number: A26655; MUID:87092266; PMID:3540939
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAR>
A;Cross-references: GB:M14628; GB:M11938; NID:g167834; PIDN:AAA33227.1; PID:g167835
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788; PMID:3901008
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116 <DEL>
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A;Reference number: S00250; MUID:88112226; PMID:2828113
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1734-1893 <WAG>
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos
F;1-818/Domain: globular head <HED>
F;89-747/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

Query Match 7.8%; Score 186.5; DB 2; Length 2116;
Best Local Similarity 21.6%; Pred. No. 0.29;
Matches 108; Conservative 86; Mismatches 205; Indels 101; Gaps 17;

```
Qy         26 SQDKS--LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE 83
              ::|||  |  ::   |     | | | |     ||  |  |: : |      :|
Db       1023 TKDKSELLRQKKKLEEELKQVQEALAAETAAKLAQEAANKKLQGEYTEL-NEKFNSEVTA 1081

Qy         84 KEKIEKERQSIRSS--PLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQL 141
               :  :|| :::: |    ::|:|: |     |||   |  :  |  ||   ::  | |
Db       1082 RSNVEKSKKTLESQLVAVNNELDEE----KKNR---DALEKKKKALDAMLEEMKDQLEST 1134

Qy        142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLI 201
              |   :  |:      :: :|:|       :|:| :     |   ||| ||| :  :|
Db       1135 GGEKKSLYDL------KVKQESDMEALRNQISELQSTIAKLEKIKSTLEGEVARLQGELE 1188

Qy        202 SKE--ANNYEEDPNKPTSWTENQAGKIPEKVT-------------PMAAIQDGLAKGEN 245
              :::   :| |:   |    |::: :: |:                :: :|  |::  |
Db       1189 AEQLAKSNVEKQKKKVELDLEDKSAQLAEETAAKQALDKLKKKLEQELSEVQTQLSEANN 1248

Qy        246 D-----------ETVSNTLTLTNGLERRTKTYSEDN----FEELQYFPNFYALLKSIDSE 290
                          ||  | | |    |:: |   |        ||::      : : :: |
Db       1249 KNVNSDSTNKHLETSFNNLKLELEAEQKAKQALEKKRLGLESELKH------VNEQLEEE 1302

Qy        291 KEAKEK--------ETLITIMKTLIDFVKMMVKYGTISPEEGVSYLE------------ 329
              |: ||          |  :: :|  |:       |  | :  :  | |:
Db       1303 KKQKESNEKRKVDLEKEVSELKDQIEEEVASKKAVTEAKNKKESELDEIKRQYADVVSSR 1362

Qy        330 --NLDEMIALQTKN-KLEKNATDNISKLFPAPSEKSHEETD------STKEEAAKMEKEY 380
                :::::  || || :|   | :   :|  |   |  | |       : :|| ||  |
Db       1363 DKSVEQLKTLQAKNEELRNTAEEAEGQLDRAERSKKKAEFDLEEAVKNLEEETAKKVKAE 1422

Qy        381 GSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFI 440
              ::| :  |  |       |     :| |::  | | |                ||::|  :
Db       1423 KAMKKAETDYRSTKSELDDAKNVSSEQYVQIKRLNEE--------------LSELRSVL 1467

Qy        441 NKQADAYVEKGILDKEEAEA 460
              ::||      |  |: ||:
Db       1468 -EEADERCNSAIKAKKTAES 1486
```

RESULT 5
B71609
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71609
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-665 <GAR>
A;Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AAC71925.1; PID:g384524
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0680w

Query Match 7.7%; Score 184; DB 2; Length 665;
Best Local Similarity 22.0%; Pred. No. 0.089;
Matches 104; Conservative 76; Mismatches 182; Indels 110; Gaps 17;

```
Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEK 84
              | || |  | |    :: : | : | || | ::    ::|  ::  |           :
Db       160 GKQDISNSNAE--NKKDVKEGVKELEEKKKEEKISDDHKVEENKKS-----------DD 205

Qy         85 EKIEKERQSIRSSPLDNKLN----VEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQ 140
               |:|: ::|       :|| :    :|:|   :  :  :: |  :   ::| :|:    :
Db       206 HKVEENKKSDDHKVEENKKSDDHKIEEVKKVEEHEEDEEEDKKEKKSENKNKDENKDEND 265

Qy        141 LDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKL 200
              |   :: || |          |||    ||             |:    ||:|  |:::|
Db       266 EDNDEISDEDEVDDDVEEDKNENDDIDDDK----------KETDKTHLEEEENEIIEKE 314

Qy        201 IS-KEANNYEEDPNKPTSW-TENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNG 258
              | |: |  :|  |  |  || :  |  ||           | :: |   :
Db       315 FSDKKKNGKNKDTKKEKSKDTEKEKSKDIEK----------EKSKDKEKEKS------- 356

Qy        259 LERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGT 318
               : :  |  :|  :|           || | ||| |||:  |   |:  |  |   |
Db       357 -KDKEKEKGKDKEKE----------KSKDIEKE-KEKDKDIEKEKSK-DTAKEKEKDKD 402

Qy        319 ISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSH----------E 364
              |  |:          :|  |: |   ||  ||  |      ::|                |
Db       403 IEKEKS-------KDMEKLKNKQNDEKKKDDNEKK----KNDKQDIHDDNDDENDMEEIE 451

Qy        365 ETDSTKEEAAKME----------KEYGSLKDSTK-DDNSNPGGKTDEPKGKTEAYLEAI 412
              | |  ::|   ||          | |:   |   ::| |  |  :| | ::|   |
Db       452 ENDDEEDEDEDMENKKKKKKGKNGNENGNENGSENGNENGNENGNENENKNESENENENE 511

Qy        413 RKNIEWLKKHDKKGNKEDYDLSKMRDF----------INKQADAYVEKGILD 454
               :|    :  ::| |::| :: :: :           ||| ::  :  |  :|
Db       512 NENENGNENENEKENEKDKNIKEIENVTNANKENYEKINKNSEITITKSNID 563
```

RESULT 6
S67593
transport protein USO1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D2552; protein YDL058w
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C;Accession: S67593; A38455; S30782
R;Bloecker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67587
A;Accession: S67593
A;Molecule type: DNA
A;Residues: 1-1790 <BLO>
A;Cross-references: EMBL:Z74106; NID:g1431058; PID:e253003; PID:g1431059; MIPS:YDL058w
A;Experimental source: strain S288C
R;Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
J. Cell Biol. 113, 245-260, 1991
A;Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transpo
A;Reference number: A38455; MUID:91185402; PMID:2010462

A;Accession: A38455
A;Molecule type: DNA
A;Residues: 1-389,'TA',392-724,'S',726-1790 <NAK>
A;Cross-references: GB:X54378; NID:g4777; PIDN:CAA38253.1; PID:g4778
A;Note: the authors translated the codon ACT for residue 768 as Ile
R;Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A;Description: An integrin analogue in Saccharomyces cerevisiae.
A;Reference number: S30782
A;Accession: S30782
A;Molecule type: DNA
A;Residues: 71-846,'E',848-923,'K',925-1252,'I',1254-1318,'V',1320-1460,'S',1462-1580,'S
A;Cross-references: EMBL:L03188
C;Genetics:
A;Gene: SGD:USO1; INT1
A;Cross-references: SGD:S0002216; MIPS:YDL058w
A;Map position: 4L
C;Keywords: coiled coil; transmembrane protein
F;326-342/Domain: transmembrane #status predicted <TM1>
F;394-410/Domain: transmembrane #status predicted <TM2>
F;617-633/Domain: transmembrane #status predicted <TM3>

Query Match 7.6%; Score 183.5; DB 2; Length 1790;
Best Local Similarity 20.3%; Pred. No. 0.33;
Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;

```
Qy      29 KSLHN--RELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI----T 82
           |||| |  ::: ||   || : :| |:                | | :   ||   |
Db     965 KSLANNYKDMQAE---NESLIKAVEE------------SKNESSIQLSNLQNKIDSMSQ 1008

Qy      83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI------DDYDSTKSGLDHKFQD--- 133
           |||  : || ||  :   |  : |:: ||   :        |:|:|  | |  | :
Db    1009 EKENFQIERGSIEKNIEQLKKTISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATT 1068

Qy     134 -DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH----- 187
            : : ::::     | |::  ::||  |:     :  |: :    |  : |:: |
Db    1069 ANDENVNKISELTKTREELEAELAA--YKNLKNELETKLETSEKALKEVKENEEHLKEEK 1126

Qy     188 -TLEDEVAEVLQKLISKEAN---------------NYEED-PNKPTSWTENQAGKIPEK 229
            || |  |  |:| |  ||                  |||   ||   : | :  :: ::
Db    1127 IQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKERQYNE-EISQLNDE 1185

Qy     230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRT----------------KTYSEDN-- 270
           :|      : : | :|||       : :  | ::                |  :| |
Db    1186 ITSTQQENESIKK-KNDELEGEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEA 1244

Qy     271 ---------------FEELQYFPNF--------------------YALLKSIDSEKEAK 294
                         :|||   ||                    |  |:  :|||  :
Db    1245 SLLESIKSVESETVKIKELQDECNFKEKEVSELEDKLKASEDKNSKYLELQK-ESEKIKE 1303

Qy     295 EKETLITIMKTLIDFVKMMVK------------------------------------ 315
           | :   | :|  :: :  : |
Db    1304 ELDAKTTELKIQLEKITNLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQ 1363

Qy     316 ------------YGTISPE--EGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEK 361
                       ||: |  | :: ||  ||:| || :|:|:    ||          |
Db    1364 AFEKERKLLNEGSSTITQEYSEKINTLE--DELIRLQNENELKAKEIDNTRSELEKVSLS 1421

Qy     362 SHEETDSTKEEAAKMEKEYGSLKDS-TKDDNSNPGGKTDEP------KGKTEAYLEAIRK 414
           : |  :  :     :: | | ||  |::|      : |       | :  |  |:  |
Db    1422 NDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAK 1481

Qy     415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILD-KEEAEAIKR 463
             | ||| ::: :||  :| | :: : |: :: :|    : |    | |::
Db    1482 VEEGLKKLEEESSKEKAELEKSKEMM-KKLESTIESNETELKSSMETIRK 1530
```

RESULT 7
A56577
microtubule-associated protein MAP 1B - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C;Accession: A56577

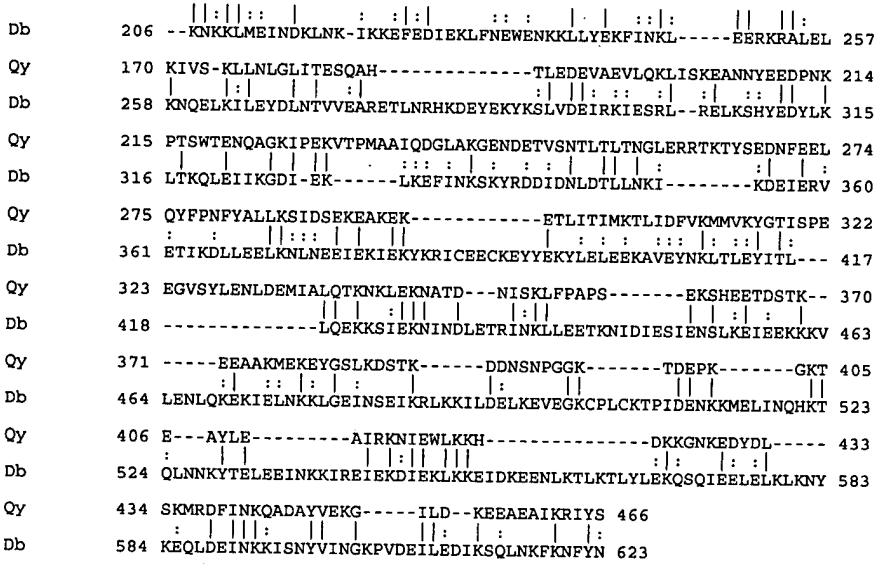
R;Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
Eur. J. Cell Biol. 57, 66-74, 1992
A;Title: Identification of two distinct microtubule binding domains on recombinant rat MA
A;Reference number: A56577; MUID:92347374; PMID:1639092
A;Accession: A56577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2364 <ZAU>
A;Cross-references: GB:X60550
A;Experimental source: brain
A;Note: nucleotide sequence not given; conceptual translation not complete
C;Superfamily: microtubule-associated protein MAP1B

Query Match 7.5%; Score 180.5; DB 2; Length 2364;
Best Local Similarity 22.7%; Pred. No. 0.65;
Matches 120; Conservative 77; Mismatches 200; Indels 131; Gaps 22;

```
Qy      22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIK-KTYPPENKPGQSNYSFVDNLNLLKA 80
           ||  |:     ::| : |      |: :  : : | |    ::|||:        :   :
Db     454 KPLSSKSVRKESKEEAPEATKASQVEKTPKVESKEKVIVKKDKPGK--------VESKPS 505

Qy      81 ITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDP--DGL 138
           :||||   || ||   : :  |  |          |::     ||   : | :: |  :
Db     506 VTEKEVPSKEEQSPVKAEVAEKAATESKPKVTKDKVVKKEIKTKP--EEKKEEKPKKEVA 563

Qy     139 HQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHT-LEDEVAEVL 197
           : | |||  ::   |      |  :    | : |     |  | :  | ||:|   ||
Db     564 KKEDKTPLKKDEKPKK-------EEAKKEIKKEIKKEEKKELKKEVKKETPLKDAKKEV- 615

Qy     198 QKLISKEANNYEEDPNK-------------PTSWTENQAG--------KIPEKVTPMAA- 235
           :|   ||     |::| |              | | |:  |          : | |  |:||
Db     616 KKDEKKEVKKEEKEPKKEIKKISKDIKKSTPLSDTKKPAALKPKVAKKEEPTKKEPIAAG 675

Qy     236 -IQD-GLAK-----GENDETVSNTL-----------------------TLTNGLERRTK 264
            ::| |  |      |:  |  :  :                      :| :  | ||
Db     676 KLKDKGKVKVIKKEGKTTEAAATAVGTAAVAAAAGVAASGPAKELEAERSLMSSPEDLTK 735

Qy     265 TYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KETLITIMKTLIDFVKMMV 314
           : |   ||:      :    |: |: |:: ||          ||| ::
Db     736 DFEELKAEEIDVAKDIKPQLELIEDEEKLKETEPGEAYVIQKETEVS------------- 782

Qy     315 KYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL------FPAPSEKSHEE 365
           |    ||:||::  |   |    ||  :|   ||   |:| |       |   ||    |
Db     783 KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFEDEGAGFEESSEAGDYE 840

Qy     366 TDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEA--YLEAIRKNI------- 416
           :  ||| : |::    :|:      |      ||   | ||   :::  |:::
Db     841 EKAETEEAEEPEEDG---EDNVSGSASKHSPTEDEEIAKAEADVHIKEKRESVASGDDRA 897

Qy     417 -----EWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE 459
                | |:| : : ::|: :  :     :|  ||  |    || |||
Db     898 EEDMDEALEKGEAEQSEEEGEEEE-----DKAEDAREEDHEPDKTEAE 940
```

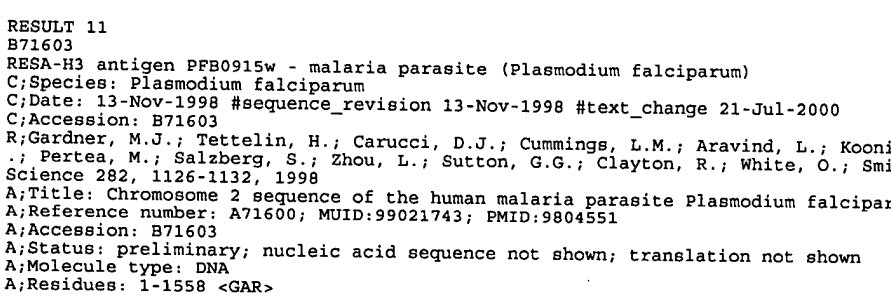
RESULT 8
A45605
mature-parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: A45605; A54517
R;Coppel, R.L.
Mol. Biochem. Parasitol. 50, 335-347, 1992
A;Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human ery
A;Reference number: A45605; MUID:92158014; PMID:1741020
A;Accession: A45605
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <COP>
A;Experimental source: Papua New Guinean isolate FC27
A;Note: sequence extracted from NCBI backbone (NCBIN:83648, NCBIP:83656)
R;Coppel, R.L.; Culvenor, J.G.; Bianco, A.E.; Crewther, P.E.; Stahl, H.D.; Brown, G.V.; A
Mol. Biochem. Parasitol. 20, 265-277, 1986
A;Title: Variable antigen associated with the surface of erythrocytes infected with matu

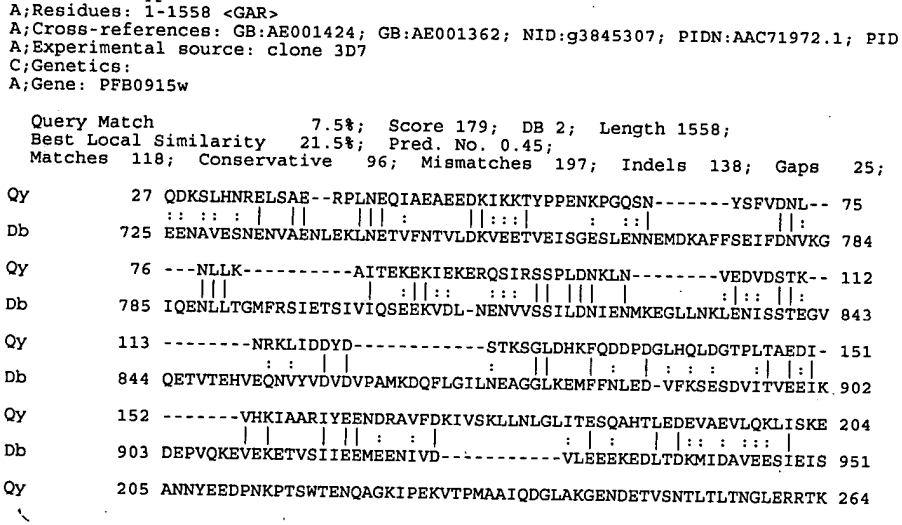
```
A;Reference number: A54517; MUID:87014571; PMID:3531849
A;Accession: A54517
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 222-443,'K',445 <CO2>
A;Cross-references: GB:M15319; NID:g160060; PID:g552170
C;Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology
C;Keywords: surface antigen; tandem repeat

  Query Match             7.5%;  Score 179.5;  DB 2;  Length 1526;
  Best Local Similarity   18.7%;  Pred. No. 0.42;
  Matches  107;  Conservative  101;  Mismatches  182;  Indels  181;  Gaps   24;

Qy         25 GSQDKSLHNRELS--AERPLNEQIAEAEEDKIKKTYPPENKPGQSN---------- 68
Db        400 GASENSEDPKKLTEQEENGTKESSEETKDDK-----PEENEKKADNKKKSKKKKKSFFQM 454

Qy         69 --YSFVDNLNL-----LKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYD 121
Db        455 LGCNFLCNKNIETDDEEETLVVKDDAKKKHKFLREA------NTEKNDNEKKDKLLGEGD 508

Qy        122 ----------------STKSGLDHKFQDDPDGL----------------HQLDGTPL 146
Db        509 KEDVKEKNDEQKDKVLGEGDKEDVKEKNDEQKDKVLGEGDKEDVKEKNDGKKDKVIGSEK 568

Qy        147 TAEDIVHKIAARI-----------YEENDRAVFDKIVSKLLNLGLITESQAHTLEDEV-- 193
Db        569 TQKEIKEKVEKRVKKKCKKKVKKGIKENDTEGNDKVKGPEIIIEEVKEEIKKQVEDGIKE 628

Qy        194 -----------AEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTP------MAAI 236
Db        629 NDTEGNDKVKGPEIITEEVKEEIKKQVEDGIKENDTEGNDKVKGPEIITEEVKEEIKKQV 688

Qy        237 QDGLAKG--ENDETVSNTLTLTNGLERRTK-----------TYSEDNF-------EELQY 276
Db        689 EEGIKENDTEGNDKVKGPEIITEEVKEEIKKQVEEGVKENDTESKDKVIGQEIITEEVK- 747

Qy        277 FPNFYALLKSIDSEKEAKEKETLITI------MKTLIDFVKMMVKYGTISPEEGV--SYL 328
Db        748 --------KEIEKQEEKGNKENILEIKDIVIGQEVIIEEVKKVIKK---KVEKGIKENHT 796

Qy        329 ENLD-----EMIALQTKNKLEKNATDNISKLFPAPSEK----------SHEETDSTKEEA 373
Db        797 ESKDKVIGQEIIVEEVKEEIEKQVEEGIKENDTESKDKVIGEEVIKGDVNEEGPENKDKV 856

Qy        374 AKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL 433
Db        857 TKQEK----VKEVKK----------EVKKK-------VKKRV---KKRNNKNERKDNVI 891

Qy        434 SK--MRDFINKQADAYVEKGILDKEEAEAIK 462
Db        892 GKEIMKEDVNEKDTANKDKEIEQEKEKEEVK 922


RESULT 9
T51505
hypothetical protein F5E19_70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51505
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-853 <SAT>
A;Cross-references: EMBL:AL391147
A;Experimental source: cultivar Columbia; BAC clone F5E19
C;Genetics:
A;Map position: 5
A;Introns: 6/2; 79/3
A;Note: F5E19_70

  Query Match             7.5%;  Score 179;  DB 2;  Length 853;
  Best Local Similarity   20.9%;  Pred. No. 0.21;
  Matches  102;  Conservative   81;  Mismatches  206;  Indels  100;  Gaps   19;

Qy         16 PIQAFPKPGGSQDKSLHNRELSAER-----PLNEQIAEAEEDKIKKTYPPEN-------- 62
Db         45 PSTTTPHSRLSLDRSSPNSKSSVERRSPKLPTPPEKSQARVAAVKGTESPQTTTRLSQIK 104

Qy         63 ---KPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDD 119
Db        105 EDLKKANERISSLEK-DKAKALDELKQAKKEAEQV-TLKLDDALKAQ--KHVEENSEIEK 160

Qy        120 YDSTKSGLDHKFQDDPDGL----------HQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
Db        161 FQAVEAGIE-AVQNNEEELKKELETVKNQHASDSAALVA---VRQELEKINEEL-AAAFD 215

Qy        170 KIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEK 229
Db        216 -AKSKAL-------SQA----------------------EDASKTAEIHAEKVDILSSE 244

Qy        230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDS 289
Db        245 LTRLKALLD----------------------STREKTAISDNEMVAKLEDEIVVLKRDLES 283

Qy        290 ----EKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEK 345
Db        284 ARGFEAEVKEKEMIVEKLNVDLEAAKMAESNAHSLSNEWQSKAKELEEQ--LEEANKLER 341

Qy        346 NAT---DNISKLFPAPSEKSHE-ETDST--KEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
Db        342 SASVSLESVMKQLEGSNDKLHDTETEITDLKERIVTLETTVAKQKEDLEVSEQRL-GSVE 400

Qy        400 EPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE 459
Db        401 EEVSKNEKEVEKLKSELETVKEEKNRALKKEQDATSRVQRLSEEKSKLLSDLESSKEEEE 460

Qy        460 AIKRIYSSL 468
Db        461 KSKKAMESL 469


RESULT 10
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: A64465
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: A64465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1005 <BUL>
A;Cross-references: GB:U67572; GB:L77117; NID:g1591958; PIDN:AAB99331.1; PID:g1591962; TI
C;Genetics:
A;Map position: REV1273394-1270377
C;Superfamily: hypothetical protein MJ1322

  Query Match             7.5%;  Score 179;  DB 2;  Length 1005;
  Best Local Similarity   22.3%;  Pred. No. 0.26;
  Matches  116;  Conservative   80;  Mismatches  164;  Indels  160;  Gaps   27;

Qy         52 DKIKKTYPPENKPGQSNYSFVDNLNLLKA-ITEKEKIEKERQSIRSSPLDNKLN-VEDVD 109
Db        159 DEFEKCY---QKMGEIVKEYEKRLERIEGELNYKENYEKE--------LKNKMSQLEE-- 205

Qy        110 STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
```

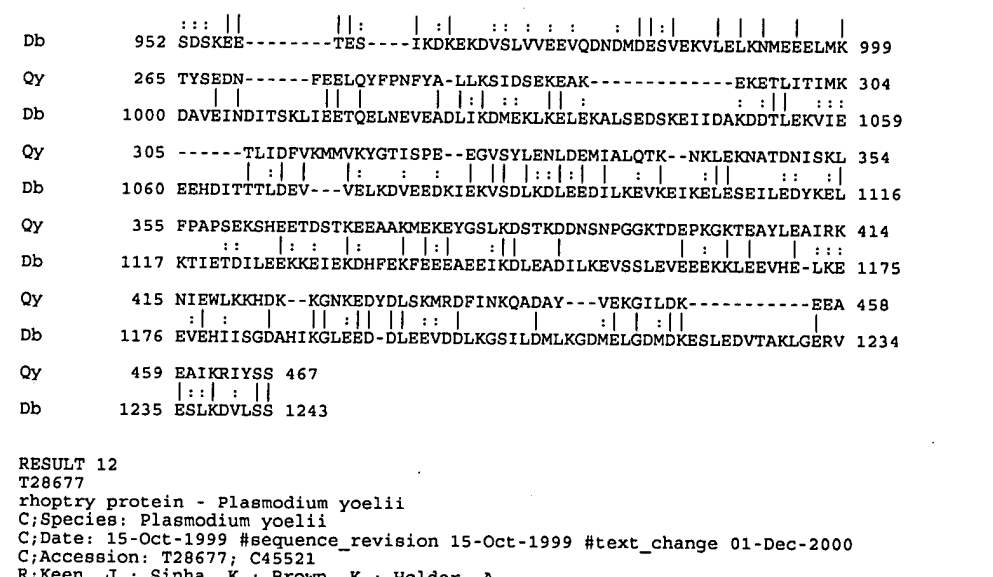


RESULT 11
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71603
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71603
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1558 <GAR>
A;Cross-references: GB:AE001424; GB:AE001362; NID:g3845307; PIDN:AAC71972.1; PID:g384530
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0915w

Query Match 7.5%; Score 179; DB 2; Length 1558;
Best Local Similarity 21.5%; Pred. No. 0.45;
Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;

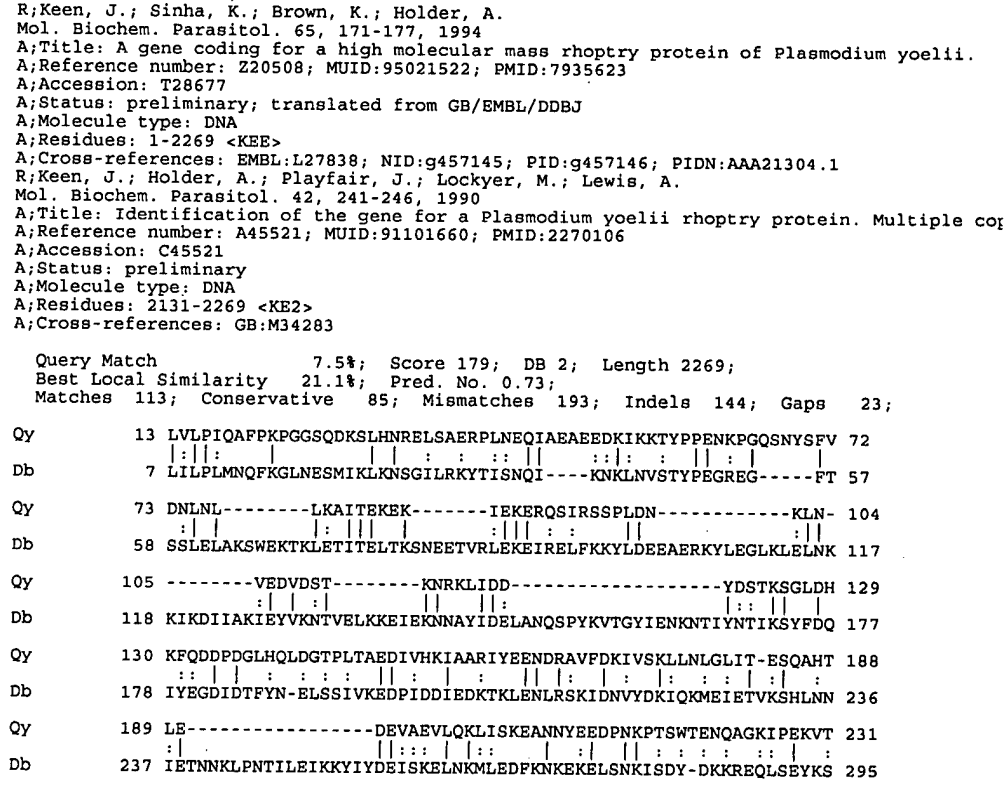






RESULT 12
T28677
rhoptry protein - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28677; C45521
R;Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A;Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A;Reference number: Z20508; MUID:95021522; PMID:7935623
A;Accession: T28677
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2269 <KEE>
A;Cross-references: EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304.1
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: C45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2131-2269 <KE2>
A;Cross-references: GB:M34283

Query Match 7.5%; Score 179; DB 2; Length 2269;
Best Local Similarity 21.1%; Pred. No. 0.73;
Matches 113; Conservative 85; Mismatches 193; Indels 144; Gaps 23;



```
Qy        232 PMAAIQDGLAKGENDET-VSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSE 290
Db        296 KMLEIRNHY----NSQTNVDNTKE-----EEAKQNYDKSN-EHMTTIPTNEDEISKIISE 345

Qy        291 KEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDN 350
Db        346 VKTM-KDEILSKVNTYIDFNK---KYKETVNSEHSQFTELTDKIKAEVSDKELKK----- 396

Qy        351 ISKLFPAPSEKSHEETDS-TKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYL 409
Db        397 --------CEQSFNDNKSLINETKNSIEKEY---------QNINTLKKVDE-------YI 432

Qy        410 EAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRI 464
Db        433 KVCKSTKESITKFSSK-------QTILKDMLNQNIKTVKETNSIDKSYIEKFEQI 480


RESULT 13
A43336
microtubule-vesicle linker CLIP-170 - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C;Accession: A43336
R;Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A;Title: CLIP-170 links endocytic vesicles to microtubules.
A;Reference number: A43336; MUID:92405160; PMID:1356075
A;Accession: A43336
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1392 <PIE>
A;Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

  Query Match             7.4%;  Score 178.5;  DB 2;  Length 1392;
  Best Local Similarity   22.4%;  Pred. No. 0.42;
  Matches  111;  Conservative   79;  Mismatches  182;  Indels  123;  Gaps   20;

Qy         34 RELSAE-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL--------NLLKA 80
Db        734 RKASSEGKSEMKKLRQQL-EAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTNLQEN 792

Qy         81 ITE----KEKIEKERQ--------------SIRSSPLD--NKL-------NVEDVDSTKN 113
Db        793 LSEVSQVKETLEKELQILKEKFAEASEEAVSVQRSMQETVNKLHQKEEQFNMLSSDLEKL 852

Qy        114 RKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS 173
Db        853 RENLAD-------MEAKFREKDEREEQL-------------IKAKEKLEND-------IA 885

Qy        174 KLLNLGLITESQAHTLEDEV------AEVLQKLISKEANNYE------EDPNKPTSWTEN 221
Db        886 EIMKMSGDNSSQLTKMNDELRLKERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQ 945

Qy        222 QAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFY 281
Db        946 EAAKKHEE-----------EKKELERKLSD-------LEKKMET-SHNQCQELK------ 980

Qy        282 ALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKN 341
Db        981 ARYERATSETKTKHEEILQNLQKTLLDTEDKLKG----AREENSGLLQELEELRKQAEKA 1036

Qy        342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
Db       1037 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNAKLQNELDTLKENNL-KNVEELNKSK 1095

Qy        400 EPKGKTEAYLEAIRKNIEWLKKH-----------DKKGNKEDYDLSKMRDFINKQADAYV 448
Db       1096 ELLTVENQKMEEFRKEIETLKQAAAQKSQQLSALQEENVKLAEELGRSRDEVTSHQKLEE 1155

Qy        449 EKGILDKEEAEAIKR 463
Db       1156 ERSVLNNQLLEMKKR 1170


RESULT 14
S22695
restin - human
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999
C;Accession: S22695; S19853
R;Bilbe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;
EMBO J. 11, 2103-2113, 1992
A;Title: Restin: a novel intermediate filament-associated protein highly expressed in the
A;Reference number: S22695; MUID:92289675; PMID:1600942
A;Accession: S22695
A;Molecule type: mRNA
A;Residues: 1-1427 <BIL>
A;Cross-references: EMBL:X64838; NID:g35998; PIDN:CAA46050.1; PID:g35999
C;Keywords: cytoskeleton

  Query Match             7.4%;  Score 178.5;  DB 2;  Length 1427;
  Best Local Similarity   22.4%;  Pred. No. 0.43;
  Matches  111;  Conservative   79;  Mismatches  182;  Indels  123;  Gaps   20;

Qy         34 RELSAE-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL--------NLLKA 80
Db        769 RKASSEGKSEMKKLRQQL-EAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTNLQEN 827

Qy         81 ITE----KEKIEKERQ--------------SIRSSPLD--NKL-------NVEDVDSTKN 113
Db        828 LSEVSQVKETLEKELQILKEKFAEASEEAVSVQRSMQETVNKLHQKEEQFNMLSSDLEKL 887

Qy        114 RKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS 173
Db        888 RENLAD-------MEAKFREKDEREEQL-------------IKAKEKLEND-------IA 920

Qy        174 KLLNLGLITESQAHTLEDEV------AEVLQKLISKEANNYE------EDPNKPTSWTEN 221
Db        921 EIMKMSGDNSSQLTKMNDELRLKERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQ 980

Qy        222 QAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFY 281
Db        981 EAAKKHEE-----------EKKELERKLSD-------LEKKMET-SHNQCQELK------ 1015

Qy        282 ALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKN 341
Db       1016 ARYERATSETKTKHEEILQNLQKTLLDTEDKLKG----AREENSGLLQELEELRKQADKA 1071

Qy        342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
Db       1072 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNAKLQNELDTLKENNL-KNVEELNKSK 1130

Qy        400 EPKGKTEAYLEAIRKNIEWLKKH-----------DKKGNKEDYDLSKMRDFINKQADAYV 448
Db       1131 ELLTVENQKMEEFRKEIETLKQAAAQKSQQLSALQEENVKLAEELGRSRDEVTSHQKLEE 1190

Qy        449 EKGILDKEEAEAIKR 463
Db       1191 ERSVLNNQLLEMKKR 1205


RESULT 15
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96796
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
```

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1871 <STO>
A;Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141
C;Genetics:
A;Gene: F28O16.15
A;Map position: 1

Query Match 7.4%; Score 177; DB 2; Length 1871;
Best Local Similarity 20.7%; Pred. No. 0.71;
Matches 116; Conservative 85; Mismatches 212; Indels 148; Gaps 22;

```
Qy        12 VLVLPIQAFPKP--GGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNY 69
             |||   : :||    | :| : |  |   |  : :     ||| ||    | :
Db      1052 VLVEEEETYPKDKHTGGEDHNDHKEEEQKENVIAKAELNTEEDSFKKVEEIEKQ------ 1105

Qy        70 SFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKN---RKLIDDYDSTKSG 126
               |:  | ::: : :: | | :  ::  ::    ||    ||:    || : |   |
Db      1106 ---DHGELKRSMVQAKRQETEEKD-KTRAMEKNETVERRKQTKDGSLGKLREGEDPELGG 1161

Qy       127 LDHKFQDDPDGLHQLDGTPLT--AEDIVHKIAARIYEENDRAVFD----KIVSKLLNLGL 180
              |: : : | : :|  | ::   | :  |   |   |    |    :  ||  :
Db      1162 --HERRGEEDRIEELVETEISDHKEKVKKKDEDYILRSQDTGKVDLGERERRSKQRKIHK 1219

Qy       181 ITESQAHTLEDEVAEVLQKLISKEAN-------NYEEDPNKPTSWTENQAGKIPEKVTP- 232
             | :   ||| ||    ::|:  |          ||:  |       :  ||||   |
Db      1220 SVEDEIGDQEDEDAEEAAAVVSRNENGSSRKVQTIEEESEK-----HKEQNKIPETSNPE 1274

Qy       233 -----------------MAAIQDGLAKGEN----------DETVSNTLTLTNGLERRTKT 265
                             | :|:   | ||          :|    :|  | | :| ||
Db      1275 VNEEDEERVVEKETKEVEAHVQELEGKTENCKDDDGEGRREERGKQGMTAENMLRQRFKT 1334

Qy       266 YSEDNFEELQYFPNFYALLKSID--SEKEAKEKETLITIMKTLIDFVKMMVKYGTI---- 319
              |:|             ::: |    |:|  ||::     ::    ||:: : |::
Db      1335 KSDD-----------GIVRKIQETKEEEPDEKKS----QESSSHVVKLVAEDGSLRNGL 1378

Qy       320 ---SPEEGVSYLENLDEMIALQTKNKLEKNATD--NISKLFPAPSEKSHEETDST----- 369
                |  || :  |||    :   |: |  :  |   :    ::|: ||
Db      1379 EFSEKESTVSKMLKLDESKEKEEHKKIRKPTEERSNAPVIEKQGNKKNAEEEMQDKIDRR 1438

Qy       370 -KEEAAKMEKEYGSLKD--------------------------STKDDNSNPGGKTD-- 399
              | :  | :: || |::                         |||:   |   ::
Db      1439 GKNQEIKGQEPYGVLRNGEHDKITEYHRGEEKGTAENVSSTKIQQTKDELEKPRKPSEIS 1498

Qy       400 -----------------EPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLS-KMRDFIN 441
                             | ||  :|   | :  |: :   :  |  ||:| :| ::|: :
Db      1499 ENHNIHEFMDSSQSQDIEEKGSDQAEKYAKQNKIQEVMNDEDK--KEEYHISERVRNEMA 1556

Qy       442 K---QADAYVEKGILDKEEAE 459
              |   | ::    |   | | |
Db      1557 KRILQVESKANDGSSKKNETE 1577
```

Search completed: January 2, 2004, 12:36:54
Job time : 22 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 18 Seconds
(without alignments)
1222.694 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2395	99.8	468	1	SG3_HUMAN	Q8wxd2 homo sapien
2	2102.5	87.6	471	1	SG3_MOUSE	P47867 mus musculu
3	2096.5	87.4	471	1	SG3_RAT	P47868 rattus norv
4	191	8.0	2464	1	MAPB_MOUSE	P14873 mus musculu
5	186.5	7.8	2116	1	MYS2_DICDI	P08799 dictyosteli
6	183.5	7.6	1790	1	USO1_YEAST	P25386 saccharomyc
7	180.5	7.5	2459	1	MAPB_RAT	P15205 rattus norv
8	179	7.5	1005	1	RA50_METJA	Q58718 methanococc
9	178.5	7.4	1427	1	REST_HUMAN	P30622 homo sapien
10	178	7.4	882	1	RA50_PYRFU	P58301 pyrococcus
11	172.5	7.2	1164	1	BAG_STRAG	P27951 streptococc
12	169	7.0	700	1	TRDN_CANFA	P82179 canis famil
13	167.5	7.0	957	1	KF5C_HUMAN	O60282 homo sapien
14	166.5	6.9	1875	1	MLP1_YEAST	Q02455 saccharomyc
15	166	6.9	2663	1	CENE_HUMAN	Q02224 homo sapien
16	166	6.9	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
17	163.5	6.8	956	1	KF5C_MOUSE	P28738 mus musculu
18	163	6.8	1251	1	RBP2_PLAVB	Q00799 plasmodium
19	161.5	6.7	1360	1	CING_XENLA	Q9ptd7 xenopus lae
20	161.5	6.7	1969	1	MYSA_CAEEL	P12844 caenorhabdi
21	161	6.7	852	1	RA50_THEMA	Q9x1x1 thermotoga
22	159.5	6.6	997	1	SCP1_RAT	Q03410 rattus norv
23	159.5	6.6	1978	1	MYHB_CHICK	P10587 gallus gall
24	159	6.6	1208	1	PCP1_SCHPO	Q92351 schizosacch
25	158.5	6.6	1526	1	MYS2_SCHPO	Q9usi6 schizosacch
26	158	6.6	1972	1	MYHB_RABIT	P35748 oryctolagus
27	157.5	6.6	1972	1	MYHB_HUMAN	P35749 homo sapien
28	157.5	6.6	2230	1	GOG4_HUMAN	Q13439 homo sapien
29	157	6.5	727	1	MFP1_ARATH	Q9lw85 arabidopsis
30	157	6.5	944	1	NUF1_YEAST	P32380 saccharomyc
31	157	6.5	2017	1	MYSN_DROME	Q99323 drosophila
32	156.5	6.5	1087	1	AKA9_RABIT	Q28628 oryctolagus
33	156.5	6.5	2468	1	MAPB_HUMAN	P46821 homo sapien
34	156	6.5	705	1	TRDN_RABIT	Q28820 oryctolagus
35	155.5	6.5	1130	1	YL17_CAEEL	Q11102 caenorhabdi
36	155.5	6.5	5085	1	PCLO_RAT	Q9jks6 rattus norv
37	154.5	6.4	1539	1	Y373_HUMAN	O15078 homo sapien
38	154	6.4	1433	1	REST_CHICK	O42184 gallus gall
39	154	6.4	1957	1	SPOF_SCHPO	Q10411 schizosacch
40	154	6.4	1962	1	MYSA_DROME	P05661 drosophila
41	153.5	6.4	1976	1	MYHA_RAT	Q9jlt0 rattus norv
42	153.5	6.4	2104	1	MYS3_SCHPO	O14157 schizosacch
43	152.5	6.4	2022	1	ANT1_ONCVO	P21249 onchocerca
44	152.5	6.4	8545	1	ANC1_CAEEL	Q9n4m4 caenorhabdi
45	152	6.3	667	1	CYL1_BOVIN	P35662 bos taurus

ALIGNMENTS

RESULT 1
SG3_HUMAN
ID SG3_HUMAN STANDARD; PRT; 468 AA.
AC Q8WXD2; Q96C83; Q96GE8; Q9Y6G7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Secretogranin III precursor (SgIII).
GN SCG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pituitary;
RA Song H., Peng Y., Huang Q., Dai M., Mao Y., Zhang Q., Mao M., Fu G.,
RA Luo M., Chen J., Hu R.;
RT "Human secretogranin III mRNA, complete cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=22093564; PubMed=12098761;
RA Rong Y.P., Liu F., Zeng L.C., Ma W.J., Wei D.Z., Han Z.G.;
RT "Cloning and characterization of a novel human secretory protein:
RT secretogranin III.";
RL Acta Biochim. Biophys. Sin. 34:411-417(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC secretory granules.
CC -!- TISSUE SPECIFICITY: Expressed in brain, heart, kidney, liver and
CC skeletal muscle.
CC ---

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF078851; AAD44483.1; -.
DR   EMBL; AF453583; AAL67431.1; -.
DR   EMBL; BC014539; AAH14539.1; -.
DR   EMBL; BC009511; AAH09511.1; ALT_INIT.
DR   Genew; HGNC:13707; SCG3.
KW   Signal; Cleavage on pair of basic residues; Polymorphism.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    468       SECRETOGRANIN III.
FT   VARIANT     125    125       S -> N (in dbSNP:2305710).
FT                                /FTId=VAR_013827.
FT   CONFLICT     79     79       K -> R (IN REF. 1).
FT   CONFLICT    167    167       A -> V (IN REF. 1 AND 2).
FT   CONFLICT    272    274       EEL -> RDF (IN REF. 1).
SQ   SEQUENCE   468 AA;  52977 MW;  633A87692A783808 CRC64;

  Query Match             99.8%;  Score 2395;  DB 1;  Length 468;
  Best Local Similarity   99.8%;  Pred. No. 6e-108;
  Matches  467;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAAFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 2
SG3_MOUSE
ID   SG3_MOUSE       STANDARD;       PRT;   471 AA.
AC   P47867;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Secretogranin III precursor (SgIII).
GN   SCG3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95001263; PubMed=7917832;
RA   Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA   Sutcliffe J.G.;
RT   "Primary structure of mouse secretogranin III and its absence from
RT   deficient mice.";
RL   J. Mol. Neurosci. 4:225-233(1993).
CC   -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC       secretory granules.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U02982; AAA56636.1; -.
DR   MGD; MGI:103032; Scg3.
KW   Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    471       SECRETOGRANIN III.
SQ   SEQUENCE   471 AA;  53326 MW;  FC1E9C381AFCA564 CRC64;

  Query Match             87.6%;  Score 2102.5;  DB 1;  Length 471;
  Best Local Similarity   88.3%;  Pred. No. 5.8e-94;
  Matches  416;  Conservative   17;  Mismatches   35;  Indels    3;  Gaps    1;

Qy          1 MGFLGTGTWILVLVL---PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKT 57
              |||| ||:|||||||   |||||||| |||||||||||||||||||||||||| |||||
Db          1 MGFLWTGSWILVLVLNSGPIQAFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKA 60

Qy         58 YPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI 117
              :| |:|| :|||| |||||||:|||||| :||||||||| | ||:||||| |||||||||
Db         61 FPSESKPSESNYSSVDNLNLLRAITEKETVEKERQSIRSPPFDNQLNVEDADSTKNRKLI 120

Qy        118 DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLN 177
              |:|||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db        121 DEYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLN 180

Qy        178 LGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
              |||||||||||||||| |||||||||||||  :|||| |||| |||||||||:||:|
Db        181 LGLITESQAHTLEDEVAEALQKLISKEANNYEETLDKPTSRTENQDGKIPEKVTPVAAVQ 240

Qy        238 DGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              ||    ||||||||||:|||||||  : ||:|||||||||||||| ||||||||||||
Db        241 DGFTNRENDETVSNTLTLSNGLERRTNPHREDDFEELQYFPNFYALLTSIDSEKEAKEKE 300

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||: ||||||
Db        301 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDETIALQTKNKLEKNTTDSKSKLFPA 360

Qy        358 PSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIE 417
              | ||| |||||||||||||||||||||||||||||||| ||||||  ||||||||||||||||
Db        361 PPEKSQEETDSTKEEAAKMEKEYGSLKDSTKDDNSNLGGKTDEATGKTEAYLEAIRKNIE 420

Qy        418 WLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||:||||||||||||||:||||||||||||||| |||||||||
Db        421 WLKKHNKKGNKEDYDLSKMRDFINQQADAYVEKGILDKEEANAIKRIYSSL 471


RESULT 3
SG3_RAT
ID   SG3_RAT         STANDARD;       PRT;   471 AA.
AC   P47868;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Secretogranin III precursor (SgIII) (1B1075).
GN   SCG3.
```

```
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95001263; PubMed=7917832;
RA   Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA   Sutcliffe J.G.;
RT   "Primary structure of mouse secretogranin III and its absence from
RT   deficient mice.";
RL   J. Mol. Neurosci. 4:225-233(1993).
RN   [2]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=90376160; PubMed=2204688;
RA   Ottiger H.-P., Battenberg E.F., Tsou A.-P., Bloom F.E.,
RA   Sutcliffe J.G.;
RT   "1B1075: a brain- and pituitary-specific mRNA that encodes a novel
RT   chromograNIN / SECretogranin-like component of intracellular
RT   vesicles.";
RL   J. Neurosci. 10:3135-3147(1990).
CC   -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC       secretory granules.
CC   -!- TISSUE SPECIFICITY: BRAIN- AND PITUITARY-SPECIFIC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U02983; AAA56637.1; -.
KW   Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    471       SECRETOGRANIN III.
SQ   SEQUENCE   471 AA;  53183 MW;  977B3F8885F33223 CRC64;

  Query Match             87.4%;  Score 2096.5;  DB 1;  Length 471;
  Best Local Similarity   87.9%;  Pred. No. 1.1e-93;
  Matches  414;  Conservative   18;  Mismatches   36;  Indels    3;  Gaps    1;

Qy          1 MGFLGTGTWILVLVL---PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKT 57
              |||| ||:|||||||    |||||||| ||||||||||||||||||||||||| ||||||
Db          1 MGFLWTGSWILVLVLNSGPIQAFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKT 60

Qy         58 YPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI 117
              || |:|| :||:| |||||||||||||| :|| :||||||| ||:|||:| ||||||||
Db         61 YPSESKPSESNFSSVDNLNLLKAITEKETVEKAKQSIRSSPPDNRLNVDDADSTKNRKLT 120

Qy        118 DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLN 177
              |:|||||||||| | |||||||||||||||||||||||| |||||||| ||||||||||
Db        121 DEYDSTKSGLDRKVQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLN 180

Qy        178 LGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
              |||||||||||||||| |||||||||||| | |||| |||| |||||||||:|| | |
Db        181 LGLITESQAHTLEDEVAEALQKLISKEANNYEEAPEKPTSRTENQDGKIPEKVTPVAATQ 240

Qy        238 DGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              ||    |||:|||||||:|||||||  : :|:|||||||||||| ||||||||||||
Db        241 DGFTNRENDDTVSNTLTLSNGLERRTNPHRDDDFEELQYFPNFYALLTSIDSEKEAKEKE 300

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              |||||||||||||||||||||||||||||||| |||||||||||| ||: ||||||
Db        301 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDETIALQTKNKLEKNTTDSKSKLFPA 360

Qy        358 PSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIE 417
              | |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db        361 PPEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNLGGKTDEAKGKTEAYLEAIRKNIE 420

Qy        418 WLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
                  |||||:||||||||||||||:|||||||||||||||| |||||||||
Db        421 WLKKHNKKGNKEDYDLSKMRDFINQQADAYVEKGILDKEEANAIKRIYSSL 471


RESULT 4
MAPB_MOUSE
ID   MAPB_MOUSE     STANDARD;      PRT;  2464 AA.
AC   P14873;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE   [Contains: MAP1 light chain LC1].
GN   MAP1B OR MTAP1B OR MTAP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DOMAIN.
RC   STRAIN=Swiss Webster; TISSUE=Brain;
RX   MEDLINE=90094539; PubMed=2480963;
RA   Noble M., Lewis S.A., Cowan N.J.;
RT   "The microtubule binding domain of microtubule-associated protein
RT   MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT   and tau.";
RL   J. Cell Biol. 109:3367-3376(1989).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC       OF MAP1B.
CC   -!- SIMILARITY: TO MAP1A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51396; CAA35761.1; -.
DR   PIR; S07549; QRMSP1.
DR   MGD; MGI:1306778; Mtap1b.
DR   GO; GO:0016358; P:dendrite morphogenesis; IMP.
DR   GO; GO:0001578; P:microtubule bundling; IMP.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 7.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?   2464       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1874   1890       MAP1B 1.
FT   REPEAT     1891   1907       MAP1B 2.
FT   REPEAT     1908   1924       MAP1B 3.
FT   REPEAT     1925   1941       MAP1B 4.
FT   REPEAT     1942   1958       MAP1B 5.
FT   REPEAT     1959   1975       MAP1B 6.
FT   REPEAT     1993   2009       MAP1B 7.
FT   REPEAT     2010   2026       MAP1B 8.
FT   REPEAT     2027   2043       MAP1B 9.
FT   REPEAT     2044   2060       MAP1B 10.
FT   DOMAIN      589    787       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
```

```
FT                                KKEE AND KKEI/V REPEATS).
SQ   SEQUENCE   2464 AA;  270408 MW;  FBD3DD99CFDBDA87 CRC64;

  Query Match             8.0%;  Score 191;  DB 1;  Length 2464;
  Best Local Similarity   21.3%;  Pred. No. 0.098;
  Matches  118;  Conservative   85;  Mismatches  186;  Indels  166;  Gaps   24;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEA------EEDKIKKTYP--PENKPGQSNYSFVD 73
              ||  |:     ::| : |      |: :       |:  :||  |   |:||
Db        550 KPVASKSVRKESKEETPEVTKTSQVEKTPKVESKEKVLVKKDKPVKTESKP--------- 600

Qy         74 NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQD 133
                 ::||||   || ||   : :  |  |          |::     ||  |: | ::
Db        601 ------SVTEKEVSSKEEQSPVKAEVAEKQATESKPKVTKDKVVKKEIKTK--LEEKKEE 652

Qy        134 DP--DGLHQLDGTPL------TAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQ 185
              |  : : : | |||           |::  :|   | :| :|    | | |        |:
Db        653 KPKKEVVKKEDKTPLKKDEKPRKEEVKKEIKKEIKKE-ERKELKKEVKK--------ETP 703

Qy        186 AHTLEDEVAEVLQKLISKEANNYEEDPNK-------------PTSWTENQAGKIPE---- 228
                : || :  :| : ||    |::| |              | | |:  :   |:
Db        704 LKDAKKEVKKEEKKEVKKE----EKEPKKEIKKISKDIKKSTPQSDTKKPSALKPKVAKK 759

Qy        229 ----KVTPMAA--IQD-GLAK-----GENDETVSNTL---------------------- 253
                  |  |:||  ::| | |      |:  |  :
Db        760 EESTKKEPLAAGKLKDKGKVKVIKKEGKTTEAAATAVGTAATTAAVVAAAGIAASGPVKE 819

Qy        254 -----TLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KET 298
                  :| : |  || : |  ||:    :    |: |: |:: ||           |||
Db        820 LEAERSLMSSPEDLTKDFEELKAEEIDVAKDIKPQLELIEDEEKLKETQPGEAYVIQKET 879

Qy        299 LITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL- 354
                ::              |    ||:||::  |   |    ||  :|   ||   |:| |
Db        880 EVS-------------KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFE 924

Qy        355 -----FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYL 409
                  |   ||    |  :  ||| : |:: |     |       :|    :  | : : :|
Db        925 DEGAGFEESSETGDYEEKAETEEAEEPEED-GEDNASGSASKHSPTEDDESAKAEADVHL 983

Qy        410 EAIRKNI------------EWLKK-----HDKKGNKEDY-----------DLSKMRDFIN 441
              :  |:::             : |:|      :::| :||           | ::  |::
Db        984 KEKRESVVSGDDRAEEDMDDVLEKGEAEQSEEEGEEEDKAEDAREEGYEPDKTEAEDYVM 1043

Qy        442 KQADAYVEKGILDKE 456
                ||    | |: :::
Db       1044 AVADKAAEAGVTEEQ 1058


RESULT 5
MYS2_DICDI
ID   MYS2_DICDI     STANDARD;      PRT;  2116 AA.
AC   P08799;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Myosin II heavy chain, non muscle.
GN   MHCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87092266; PubMed=3540939;
RA   Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT   "Conserved protein domains in a myosin heavy chain gene from
RT   Dictyostelium discoideum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN   [2]
RP   PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC   STRAIN=AX2;
RX   MEDLINE=90353583; PubMed=2387408;
RA   Lueck-Vielmeter D., Schleicher M., Grabatin B., Wippler J.,
RA   Gerisch G.;
RT   "Replacement of threonine residues by serine and alanine in a
RT   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL   FEBS Lett. 269:239-243(1990).
RN   [3]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=88112226; PubMed=2828113;
RA   Wagle G., Noegel A., Scheel J., Gerisch G.;
RT   "Phosphorylation of threonine residues on cloned fragments of the
RT   Dictyostelium myosin heavy chain.";
RL   FEBS Lett. 227:71-75(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345066; PubMed=7619795;
RA   Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA   Rayment I.;
RT   "X-ray structures of the myosin motor domain of Dictyostelium
RT   discoideum complexed with MgADP.BeFx and MgADP.AlF4-.";
RL   Biochemistry 34:8960-8972(1995).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345067; PubMed=7619796;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT   truncated head of Dictyostelium discoideum myosin to 2.7-A
RT   resolution.";
RL   Biochemistry 34:8973-8981(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=96206189; PubMed=8611530;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL   Biochemistry 35:5404-5417(1996).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX   MEDLINE=97452580; PubMed=9305951;
RA   Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT   "X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes
RT   of the Dictyostelium discoideum myosin motor domain.";
RL   Biochemistry 36:11619-11628(1997).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=98070605; PubMed=9405148;
RA   Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT   "X-ray crystal structure and solution fluorescence characterization
RT   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT   Dictyostelium discoideum myosin motor domain.";
RL   J. Mol. Biol. 274:394-407(1997).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC   -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC       INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC       LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC       (MLC-2).
CC   -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC       CORTEX.
CC   -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
CC       SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC       THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC       ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC       POSITION (688).
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   -!- SIMILARITY: Contains 1 IQ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M14628; AAA33227.1; -.
DR   PIR; A26655; A26655.
DR   PDB; 1MMA; 03-DEC-97.
DR   PDB; 1MMD; 17-AUG-96.
DR   PDB; 1MMG; 03-DEC-97.
DR   PDB; 1MMN; 03-DEC-97.
DR   PDB; 1MND; 17-AUG-96.
DR   PDB; 1MNE; 17-AUG-96.
DR   PDB; 1VOM; 23-DEC-96.
DR   PDB; 1LVK; 28-JAN-98.
DR   PDB; 1D0X; 20-DEC-00.
DR   PDB; 1D0Y; 20-DEC-00.
DR   PDB; 1D0Z; 20-DEC-00.
DR   PDB; 1D1A; 20-DEC-00.
DR   PDB; 1D1B; 20-DEC-00.
DR   PDB; 1D1C; 20-DEC-00.
DR   PDB; 1FMV; 20-DEC-00.
DR   PDB; 1FMW; 20-DEC-00.
DR   PDB; 1G8X; 17-JAN-01.
DR   PDB; 1JWY; 07-NOV-01.
DR   PDB; 1JX2; 07-NOV-01.
DR   DictyDb; DD01008; mhcA.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW   Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT   DOMAIN        1    761       MYOSIN HEAD-LIKE.
FT   DOMAIN      762    791       IQ.
FT   DOMAIN      817   2116       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP.
FT   DOMAIN      638    660       ACTIN-BINDING.
FT   DOMAIN      738    752       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (DI-) (POTENTIAL).
FT   MOD_RES     678    678       ALKYLATION (SH-1).
FT   MOD_RES    1823   1823       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    1833   1833       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    2029   2029       PHOSPHORYLATION (BY MHCK).
FT   TURN          3      5
FT   TURN          7      8
FT   HELIX        10     15
FT   HELIX        25     28
FT   STRAND       34     37
FT   TURN         41     42
FT   STRAND       48     55
FT   STRAND       59     63
FT   STRAND       69     73
FT   HELIX        74     76
FT   STRAND       78     79
FT   HELIX        83     85
FT   TURN         86     87
FT   STRAND       90     90
FT   HELIX        91     93
FT   HELIX        99    110
FT   TURN        111    113
FT   STRAND      116    119
FT   TURN        120    121
FT   STRAND      122    126
FT   HELIX       137    142
FT   TURN        143    145
FT   HELIX       148    150
FT   HELIX       155    169
FT   TURN        170    170
FT   STRAND      173    179
FT   TURN        181    182
FT   HELIX       185    200
FT   HELIX       210    226
FT   STRAND      227    228
FT   TURN        234    235
FT   STRAND      236    237
FT   STRAND      240    247
FT   TURN        249    250
FT   STRAND      253    261
FT   HELIX       265    268
FT   TURN        269    269
FT   TURN        273    274
FT   STRAND      278    278
FT   HELIX       279    287
FT   HELIX       290    296
FT   TURN        297    297
FT   HELIX       301    303
FT   TURN        305    307
FT   TURN        316    317
FT   HELIX       320    334
FT   TURN        335    335
FT   HELIX       338    355
FT   TURN        356    356
FT   STRAND      360    360
FT   STRAND      368    368
FT   HELIX       373    382
```

Query Match 7.8%; Score 186.5; DB 1; Length 2116;
Best Local Similarity 21.6%; Pred. No. 0.13;
Matches 108; Conservative 86; Mismatches 205; Indels 101; Gaps 17;

```
Qy         26 SQDKS--LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE 83
              ::|||  |  ::   |    | | | |   |      ||  |  |: : |      :|
Db       1023 TKDKSELLRQKKKLEEELKQVQEALAAETAAKLAQEAANKKLQGEYTEL-NEKFNSEVTA 1081

Qy         84 KEKIEKERQSIRSS--PLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQL 141
              :  :|| :::: |    ::|:|: |     |||   |  :  |  ||   ::  | |
Db       1082 RSNVEKSKKTLESQLVAVNNELDEE----KKNR---DALEKKKKALDAMLEEMKDQLEST 1134

Qy        142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLI 201
              |   :  |:       :: :|:|       :|:| :      |   ||| ||| : :|
Db       1135 GGEKKSLYDL------KVKQESDMEALRNQISELQSTIAKLEKIKSTLEGEVARLQGELE 1188

Qy        202 SKE--ANNYEEDPNKPTSWTENQAGKIPEKVT-------------PMAAIQDGLAKGEN 245
              :::   :| |:   |     |::: :: |:                 :: :|  |:  |
Db       1189 AEQLAKSNVEKQKKKVELDLEDKSAQLAEETAAKQALDKLKKKLEQELSEVQTQLSEANN 1248

Qy        246 D----------ETVSNTLTLTNGLERRTKTYSEDN----FEELQYFPNFYALLKSIDSE 290
                         ||  | | |    |:: |    |         ||::      : : :: |
Db       1249 KNVNSDSTNKHLETSFNNLKLELEAEQKAKQALEKKRLGLESELKH------VNEQLEEE 1302

Qy        291 KEAKEK--------ETLITIMKTLIDFVKMMVKYGTISPEEGVSYLE------------ 329
              |: ||          |  :: :|  |:       |  | :  :  | |:
Db       1303 KKQKESNEKRKVDLEKEVSELKDQIEEEVASKKAVTEAKNKKESELDEIKRQYADVVSSR 1362

Qy        330 --NLDEMIALQTKN-KLEKNATDNISKLFPAPSEKSHEETD------STKEEAAKMEKEY 380
                :::::  || || :|    | :   :|  |    |    | |       : :|| ||  |
Db       1363 DKSVEQLKTLQAKNEELRNTAEEAEGQLDRAERSKKKAEFDLEEAVKNLEEETAKKVKAE 1422

Qy        381 GSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFI 440
              ::| :  |  |       |     :| |::  | | |                  ||::|  :
Db       1423 KAMKKAETDYRSTKSELDDAKNVSSEQYVQIKRLNEE--------------LSELRSVL 1467

Qy        441 NKQADAYVEKGILDKEEAEA 460
              ::||      |  |: ||:
Db       1468 -EEADERCNSAIKAKKTAES 1486
```

```
RESULT 6
USO1_YEAST
ID   USO1_YEAST     STANDARD;      PRT;  1790 AA.
AC   P25386;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Intracellular protein transport protein USO1.
GN   USO1 OR INT1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180-1A;
RX   MEDLINE=91185402; PubMed=2010462;
RA   Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA   Yamasaki M.;
RT   "A cytoskeleton-related gene, uso1, is required for intracellular
RT   protein transport in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 113:245-260(1991).
RN   [2]
RP   SEQUENCE OF 782-1790 FROM N.A.
RA   Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA   Kendrick K.E.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Bai Y., Symington L.S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC       MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC       ER AND THE GOLGI COMPLEX.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC       OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC       COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC   -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54378; CAA38253.1; -.
DR   EMBL; L03188; AAB00143.1; -.
DR   EMBL; U53668; AAB66659.1; -.
DR   SGD; S0002216; USO1.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR006955; Uso1_p115_C.
DR   InterPro; IPR006953; Uso1_p115_head.
DR   Pfam; PF04871; Uso1_p115_C; 1.
DR   Pfam; PF04869; Uso1_p115_head; 1.
KW   Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT   DOMAIN        1    724       GLOBULAR HEAD.
FT   DOMAIN      725   1790       COILED COIL (POTENTIAL).
FT   DOMAIN      465    487       CHARGED (HYPER-HYDROPHILIC).
FT   DOMAIN      991   1790       DISPENSABLE FOR THE PROTEIN FUNCTION.
FT   DOMAIN     1172   1786       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT    847    847       G -> E (IN REF. 2).
FT   CONFLICT    924    924       E -> K (IN REF. 2).
FT   CONFLICT   1253   1253       V -> I (IN REF. 2).
FT   CONFLICT   1319   1319       I -> V (IN REF. 2).
FT   CONFLICT   1461   1461       N -> S (IN REF. 2).
FT   CONFLICT   1581   1581       G -> S (IN REF. 2).
FT   CONFLICT   1600   1600       I -> V (IN REF. 2).
FT   CONFLICT   1661   1661       R -> S (IN REF. 2).
FT   CONFLICT   1772   1772       D -> DEEDDEE (IN REF. 2).

SQ   SEQUENCE   1790 AA;  206424 MW;  6CE2B216E9FD4818 CRC64;

  Query Match             7.6%;  Score 183.5;  DB 1;  Length 1790;
  Best Local Similarity   20.3%;  Pred. No. 0.15;
  Matches  120;  Conservative   91;  Mismatches  200;  Indels  179;  Gaps   23;

Qy         29 KSLHN--RELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI----T 82
              |||  |  ::: ||   || : :| |:                | |  :   ||    |
Db        965 KSLANNYKDMQAE---NESLIKAVEE------------SKNESSIQLSNLQNKIDSMSQ 1008

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI------DDYDSTKSGLDHKFQD--- 133
              |||  : || ||  :    |  : |:: ||   :       |:|:|  | |  | :
Db       1009 EKENFQIERGSIEKNIEQLKKTISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATT 1068

Qy        134 -DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH----- 187
               : : ::::     | |::  ::||  |:     :  |: :    |  : |:: |
Db       1069 ANDENVNKISELTKTREELEAELAA--YKNLKNELETKLETSEKALKEVKENEEHLKEEK 1126

Qy        188 -TLEDEVAEVLQKLISKEAN---------------NYEED-PNKPTSWTENQAGKIPEK 229
               || |  |  |:| |  ||                |||   ||   : | :  :: ::
Db       1127 IQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKERQYNE-EISQLNDE 1185

Qy        230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRT----------------KTYSEDN-- 270
               :|      : : | :|||       : :  | ::              |  :| |
Db       1186 ITSTQQENESIKK-KNDELEGEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEA 1244

Qy        271 --------------FEELQYFPNF--------------------YALLKSIDSEKEAK 294
                          :|||   ||                    |  |:  :||| :
Db       1245 SLLESIKSVESETVKIKELQDECNFKEKEVSELEDKLKASEDKNSKYLELQK-ESEKIKE 1303

Qy        295 EKETLITIMKTLIDFVKMMVK------------------------------------- 315
               | :   | :|  :: :  : |
Db       1304 ELDAKTTELKIQLEKITNLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQ 1363

Qy        316 ------------YGTISPE--EGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEK 361
                         ||: |  | :: ||  ||:| || :|:|:    ||        |
Db       1364 AFEKERKLLNEGSSTITQEYSEKINTLE--DELIRLQNENELKAKEIDNTRSELEKVSLS 1421

Qy        362 SHEETDSTKEEAAKMEKEYGSLKDS-TKDDNSNPGGKTDEP------KGKTEAYLEAIRK 414
               : |  :  :     :: |  | ||  |::|      : |       | :  |  |:  |
Db       1422 NDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAK 1481

Qy        415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILD-KEEAEAIKR 463
               | ||| ::: :||  :| | :: : |: :: :|    : |   | |::
Db       1482 VEEGLKKLEEESSKEKAELEKSKEMM-KKLESTIESNETELKSSMETIRK 1530


RESULT 7
MAPB_RAT
ID   MAPB_RAT       STANDARD;      PRT;  2459 AA.
AC   P15205; Q62958; Q9ER21; Q9QW92;
DT   01-APR-1990 (Rel. 14, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE   light chain LC1].
GN   MAP1B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-142 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Testis;
RX   MEDLINE=96257242; PubMed=8666295;
RA   Liu D., Fischer I.;
RT   "Isolation and sequencing of the 5' end of the rat microtubule-
RT   associated protein (MAP1B)-encoding cDNA.";
RL   Gene 172:307-308(1996).
RN   [2]
RP   SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX   MEDLINE=92347374; PubMed=1639092;
```

```
RA   Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT   "Identification of two distinct microtubule binding domains on
RT   recombinant rat MAP 1B.";
RL   Eur. J. Cell Biol. 57:66-74(1992).
RN   [3]
RP   SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Spinal cord;
RX   MEDLINE=90059871; PubMed=2555150;
RA   Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA   Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT   "Neuraxin, a novel putative structural protein of the rat central
RT   nervous system that is immunologically related to microtubule-
RT   associated protein 5.";
RL   EMBO J. 8:2879-2888(1989).
RN   [4]
RP   DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX   MEDLINE=97405699; PubMed=9260743;
RA   Ma D., Nothias F., Boyne L.J., Fischer I.;
RT   "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT   in rat CNS and PNS during development.";
RL   J. Neurosci. Res. 49:319-332(1997).
CC   -!- FUNCTION: The function of brain MAPS is essentially unknown.
CC       Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC       that accompany neurite extension. Possibly MAP1B Binds to at least
CC       two tubulin subunits in the polymer, and this bridging of subunits
CC       might be involved in nucleating microtubule polymerization and in
CC       stabilizing microtubules.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC       cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC       heart or muscle.
CC   -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC       nerve levels are high early in development but decrease during
CC       postnatal development and are low in adults. In dorsal root
CC       ganglia levels remain high throughout development.
CC   -!- INDUCTION: By nerve growth factor.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC       from MAP1B by proteolytic processing. It is free to associate with
CC       both MAP1A and MAP1B. It interacts with the amino-terminal region
CC       of MAP1B (By similarity).
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: TO MAP1A.
CC   -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC       2459) was originally described as neuraxin in Ref.3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U52950; AAB17068.1; -.
DR   EMBL; X60370; CAC16162.1; -.
DR   EMBL; X16623; CAA34620.1; ALT_SEQ.
DR   PIR; A56577; A56577.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 8.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?    2459       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1869    1885       MAP1B 1.
FT   REPEAT     1886    1902       MAP1B 2.
FT   REPEAT     1903    1919       MAP1B 3.
FT   REPEAT     1920    1936       MAP1B 4.
FT   REPEAT     1937    1953       MAP1B 5.
FT   REPEAT     1954    1970       MAP1B 6.
FT   REPEAT     1988    2004       MAP1B 7.
FT   REPEAT     2005    2021       MAP1B 8.
FT   REPEAT     2022    2038       MAP1B 9.
FT   REPEAT     2039    2055       MAP1B 10.
FT   DOMAIN      559    1035       GLU-RICH.
FT   DOMAIN      588     786       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                 KKEE AND KKEI/V REPEATS).
FT   DOMAIN     2224    2312       LYS-RICH.
FT   CONFLICT    127     127       M -> V (IN REF. 1).
FT   CONFLICT    140     140       T -> S (IN REF. 1).
FT   CONFLICT   2112    2112       R -> K (IN REF. 3).
FT   CONFLICT   2169    2169       L -> I (IN REF. 3).
SQ   SEQUENCE   2459 AA;  269497 MW;  2E3F6872DEDB8BA2 CRC64;

  Query Match             7.5%;  Score 180.5;  DB 1;  Length 2459;
  Best Local Similarity   22.7%;  Pred. No. 0.31;
  Matches  120;  Conservative   77;  Mismatches  200;  Indels  131;  Gaps   22;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIK-KTYPPENKPGQSNYSFVDNLNLLKA 80
              ||  |:     ::| : |     |: :  : : | |    ::|||:        :   :
Db        549 KPLSSKSVRKESKEEAPEATKASQVEKTPKVESKEKVIVKKDKPGK--------VESKPS 600

Qy         81 ITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDP--DGL 138
              :||||   || ||   : :  |   |        |::     ||   : | :: |  :
Db        601 VTEKEVPSKEEQSPVKAEVAEKAATESKPKVTKDKVVKKEIKTKP--EEKKEEKPKKEVA 658

Qy        139 HQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHT-LEDEVAEVL 197
              : | |||  ::   |       | :    | : |      |  | :  | |:|   ||
Db        659 KKEDKTPLKKDEKPKK-------EEAKKEIKKEIKKEEKKELKKEVKKETPLKDAKKEV- 710

Qy        198 QKLISKEANNYEEDPNK------------PTSWTENQAG--------KIPEKVTPMAA- 235
              :|   ||    |::| |                | | |: |         : | |  |:||
Db        711 KKDEKKEVKKEEKEPKKEIKKISKDIKKSTPLSDTKKPAALKPKVAKKEEPTKKEPIAAG 770

Qy        236 -IQD-GLAK-----GENDETVSNTL-----------------------TLTNGLERRTK 264
               ::| |  |     |:  |  :  :                       :| :  |  ||
Db        771 KLKDKGKVKVIKKEGKTTEAAATAVGTAAVAAAAGVAASGPAKELEAERSLMSSPEDLTK 830

Qy        265 TYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KETLITIMKTLIDFVKMMV 314
              : |   ||:    :    |: |: |:: ||          ||| ::
Db        831 DFEELKAEEIDVAKDIKPQLELIEDEEKLKETEPGEAYVIQKETEVS------------ 877

Qy        315 KYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL------FPAPSEKSHEE 365
              |    ||:||::  |   |    ||  :|   ||   |:| |      |   ||   |
Db        878 KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFEDEGAGFEESSEAGDYE 935

Qy        366 TDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEA--YLEAIRKNI------- 416
              :  ||| : |::    :|:     |       ||   | ||  :::  |:::
Db        936 EKAETEEAEEPEEDG---EDNVSGSASKHSPTEDEEIAKAEADVHIKEKRESVASGDDRA 992

Qy        417 -----EWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE 459
                  | |:| : : ::|: :  :     :|  ||  |    || |||
Db        993 EEDMDEALEKGEAEQSEEEGEEEE-----DKAEDAREEDHEPDKTEAE 1035


RESULT 8
RA50_METJA
ID   RA50_METJA     STANDARD;      PRT;  1005 AA.
AC   Q58718;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR MJ1322.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
```

```
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67572; AAB99331.1; -.
DR   PIR; A64465; A64465.
DR   TIGR; MJ1322; -.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003405; SMC_C.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF04423; Rad50_zn_hook; 1.
DR   Pfam; PF02483; SMC_C; 1.
DR   Pfam; PF02463; SMC_N; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      32     39       ATP (BY SIMILARITY).
FT   DOMAIN      158    849       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1005 AA;  119387 MW;  9BBBB48173E788F3 CRC64;

  Query Match             7.5%;  Score 179;  DB 1;  Length 1005;
  Best Local Similarity   22.3%;  Pred. No. 0.12;
  Matches  116;  Conservative   80;  Mismatches  164;  Indels  160;  Gaps   27;

Qy         52 DKIKKTYPPENKPGQSNYSFVDNLNLLKA-ITEKEKIEKERQSIRSSPLDNKLN-VEDVD 109
              |: :| |    | |:    :   |  ::  :  ||  |||        | ||:: :|:
Db        159 DEFEKCY---QKMGEIVKEYEKRLERIEGELNYKENYEKE--------LKNKMSQLEE-- 205

Qy        110 STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
               ||:||::  |    :  :|:|     :: :   |  |  ::|:     ||  ||:
Db        206 --KNKKLMEINDKLNK-IKKEFEDIEKLFNEWENKKLLYEKFINKL-----EERKRALEL 257

Qy        170 KIVS-KLLNLGLITESQAH-------------TLEDEVAEVLQKLISKEANNYEEDPNK 214
              |    |:|   | |  :|              :| ||: ::  :|  :|  :: ||  |
Db        258 KNQELKILEYDLNTVVEARETLNRHKDEYEKYKSLVDEIRKIESRL--RELKSHYEDYLK 315

Qy        215 PTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEEL 274
              |   |   | | ||      ::: : | :  : : |  || | :          :|  | :
Db        316 LTKQLEIIKGDI-EK------LKEFINKSKYRDDIDNLDTLLNKI--------KDEIERV 360

Qy        275 QYFPNFYALLKSIDSEKEAKEK------------ETLITIMKTLIDFVKMMVKYGTISPE 322
              :   :    ||::: | |  ||            |  : : :  ::: |: ::| |:
Db        361 ETIKDLLEELKNLNEEIEKIEKYKRICEECKEYYEKYLELEEKAVEYNKLTLEYITL--- 417

Qy        323 EGVSYLENLDEMIALQTKNKLEKNATD---NISKLFPAPS-------EKSHEETDSTK-- 370
                 || |  :|||  |    |:||             | | :| :  |
Db        418 -------------LQEKKSIEKNINDLETRINKLLEETKNIDIESIENSLKEIEEKKKV 463

Qy        371 -----EEAAKMEKEYGSLKDSTK------DDNSNPGGK-------TDEPK-------GKT 405
```

```
                 :|  :: |: | :    |      |:     ||        || |        ||
Db        464 LENLQKEKIELNKKLGEINSEIKRLKKILDELKEVEGKCPLCKTPIDENKKMELINQHKT 523

Qy        406 E---AYLE---------AIRKNIEWLKKH--------------DKKGNKEDYDL----- 433
              :    | |           | |:|| |||                :|:   |: :|
Db        524 QLNNKYTELEEINKKIREIEKDIEKLKKEIDKEENLKTLKTLYLEKQSQIEELELKLKNY 583

Qy        434 SKMRDFINKQADAYVEKG-----ILD--KEEAEAIKRIYS 466
               :  | |||:   ||  |      ||:  | :    |  |:
Db        584 KEQLDEINKKISNYVINGKPVDEILEDIKSQLNKFKNFYN 623


RESULT 9
REST_HUMAN
ID   REST_HUMAN     STANDARD;      PRT;  1427 AA.
AC   P30622;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
DE   Sternberg intermediate filament associated protein).
GN   RSN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood monocytes;
RX   MEDLINE=92289675; PubMed=1600942;
RA   Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M.,
RA   Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
RA   de Wolf-Peeters C., Shipman R.;
RT   "Restin: a novel intermediate filament-associated protein highly
RT   expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL   EMBO J. 11:2103-2113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92405160; PubMed=1356075;
RA   Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT   "CLIP-170 links endocytic vesicles to microtubules.";
RL   Cell 70:887-900(1992).
CC   -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC       THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC       CYTOSKELETON.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P30622-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P30622-2; Sequence=VSP_000765;
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC       OF HODGKIN'S DISEASE.
CC   -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64838; CAA46050.1; -.
DR   EMBL; M97501; AAA35693.1; -.
DR   PIR; S22695; S22695.
DR   Genew; HGNC:10461; RSN.
DR   MIM; 179838; -.
DR   GO; GO:0005768; C:endosome; TAS.
DR   GO; GO:0005882; C:intermediate filament; TAS.
DR   GO; GO:0015630; C:microtubule cytoskeleton; TAS.
DR   GO; GO:0008017; F:microtubule binding activity; TAS.
```

```
DR   GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR   InterPro; IPR000938; CAP-Gly.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01302; CAP_GLY; 2.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 2.
DR   PROSITE; PS50245; CAP_GLY_2; 2.
KW   Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
FT   DOMAIN       78    120       CAP-GLY 1.
FT   DOMAIN      143    204       SER-RICH.
FT   DOMAIN      232    274       CAP-GLY 2.
FT   DOMAIN      304    331       SER-RICH.
FT   DOMAIN      350   1342       COILED COIL (POTENTIAL).
FT   DOMAIN     1408   1421       CCHC-BOX.
FT   VARSPLIC    457    491       Missing (in isoform Short).
FT                                /FTId=VSP_000765.
FT   CONFLICT   1069   1069       D -> E (IN REF. 2).
SQ   SEQUENCE   1427 AA;  160989 MW;  0A4F166DD94254E8 CRC64;

  Query Match             7.4%;  Score 178.5;  DB 1;  Length 1427;
  Best Local Similarity   22.4%;  Pred. No. 0.2;
  Matches  111;  Conservative   79;  Mismatches  182;  Indels  123;  Gaps   20;

Qy         34 RELSAE-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL--------NLLKA 80
              |: |:|     : | :|: || | :||     :|        |    |         || :
Db        769 RKASSEGKSEMKKLRQQL-EAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTNLQEN 827

Qy         81 ITE----KEKIEKERQ-------------SIRSSPLD--NKL-------NVEDVDSTKN 113
              ::|    || :||| |              |:: |  :  |||       |:   |  |
Db        828 LSEVSQVKETLEKELQILKEKFAEASEEAVSVQRSMQETVNKLHQKEEQFNMLSSDLEKL 887

Qy        114 RKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS 173
              |: : |       :: ||::  :   ||             | |:   |||       ::
Db        888 RENLAD-------MEAKFREKDEREEQL------------IKAKEKLEND-------IA 920

Qy        174 KLLNLGLITESQAHTLEDEV------AEVLQKLISKEANNYE------EDPNKPTSWTEN 221
              ::: :     ||   : ||:       | ||  ::|   |         ||       ::
Db        921 EIMKMSGDNSSQLTKMNDELRLKERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQ 980

Qy        222 QAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFY 281
              :| |  |:             | | :  :|:       ||:: :| | :  :||:
Db        981 EAAKKHEE-----------EKKELERKLSD-------LEKKMET-SHNQCQELK------ 1015

Qy        282 ALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKN 341
              |  :   || : | :| |  : |||:|    :      : ||     |: |:|:     |
Db       1016 ARYERATSETKTKHEEILQNLQKTLLDTEDKLKG----AREENSGLLQELEELRKQADKA 1071

Qy        342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
              |  : | |  :  :     ||:      : ||:  ||:: |  :||::    |     |:
Db       1072 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNAKLQNELDTLKENNL-KNVEELNKSK 1130

Qy        400 EPKGKTEAYLEAIRKNIEWLKKH----------DKKGNKEDYDLSKMRDFINKQADAYV 448
              |      :|  || || ||:                ::  |   :| : || :
Db       1131 ELLTVENQKMEEFRKEIETLKQAAAQKSQQLSALQEENVKLAEELGRSRDEVTSHQKLEE 1190

Qy        449 EKGILDKEEAEAIKR 463
              |: :|: :  |  ||
Db       1191 ERSVLNNQLLEMKKR 1205


RESULT 10
RA50_PYRFU
ID   RA50_PYRFU     STANDARD;      PRT;   882 AA.
AC   P58301;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR PF1167.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX   MEDLINE=20485553; PubMed=11029422;
RA   Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
RA   Carney J.P.;
RT   "Mre11 and Rad50 from Pyrococcus furiosus: cloning and biochemical
RT   characterization reveal an evolutionarily conserved multiprotein
RT   machine.";
RL   J. Bacteriol. 182:6036-6041(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX   MEDLINE=20348838; PubMed=10892749;
RA   Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
RA   Carney J.P., Tainer J.A.;
RT   "Structural biology of Rad50 ATPase: ATP-driven conformational
RT   control in DNA double-strand break repair and the ABC-ATPase
RT   superfamily.";
RL   Cell 101:789-800(2000).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site.
CC   -!- SUBUNIT: Forms a complex with mre11.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010225; AAL81291.1; -.
DR   PDB; 1F2T; 20-SEP-00.
DR   PDB; 1F2U; 02-AUG-00.
DR   PDB; 1II8; 30-MAY-01.
DR   PDB; 1L8D; 28-AUG-02.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF04423; Rad50_zn_hook; 1.
DR   Pfam; PF02463; SMC_N; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
KW   Complete proteome.
FT   NP_BIND      30     37       ATP.
FT   DOMAIN      148    744       COILED COIL (POTENTIAL).
SQ   SEQUENCE   882 AA;  103839 MW;  3ADCBD250382A99E CRC64;

  Query Match             7.4%;  Score 178;  DB 1;  Length 882;
  Best Local Similarity   22.1%;  Pred. No. 0.12;
  Matches  112;  Conservative   90;  Mismatches  172;  Indels  132;  Gaps   21;

Qy         39 ERPLNEQIAEAEE--DKIK---KTYPPENKPGQSNYSFVDNLNLLKA-----ITEKEKIE 88
              :| | |:|    ||  :|:|    |    : |    |     ||     ||       : :| |||
Db        260 KRTLEERIKNTEEYLEKLKEKEKELEEQVKEITSIKKDVDAYLALKEFKNEYLDKKYKIE 319

Qy         89 KERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTA 148
              ||    |   | |::         |  : ::: :| |  |:::
Db        320 KE--LTRVEELINEIQ-------KRIEELNEKESEKEKLENE----------------K 353

Qy        149 EDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNY 208
              ::|::|:|     | |  ::::| :|  ||  : |       ::: ::|::| :|:
```

```
Db        354 KEILNKLAIL---EKDHQLYEEIKAKKENLRQLKEKLGDKSPEDIKKLLEELETKK-TTI 409

Qy        209 EEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL--AKGE--------NDE----------- 247
              ||: |: |     : |::  |:  :    : |  |||:          ||
Db        410 EEERNEIT----QRIGELKNKIGDLKTAIEELKKAKGKCPVCGRELTDEHREELLSKYHL 465

Qy        248 -------TVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDS----------E 290
                    |::  :   : |||  :    :  : |          ::||:           |
Db        466 DLNNSKNTLAKLIDRKSELERELRRIDME-IKRLTPLLTVAEQIRSIEEELNVVNLEKIE 524

Qy        291 KEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIA--LQTKNKLEKNAT 348
              | | | | |:  ::||   :: :        | :  |  |:: :|  :  | :|||
Db        525 KNATEYEKLLEELRTLEGRIRGLA--------EDLKKLAPLEKKLAALIHKKQELEKELK 576

Qy        349 DNISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKD---------DNSNPGGKTD 399
              :  :|| :    || |: ||   |  :: | | :| :| |:             :
Db        577 ELNTKL-ESFGFPKSVEDLDSKLRELEEIYKRYLTLLNSKKELEITQREIAKAKETLEMSF 635

Qy        400 EPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDY----------------------DLSKMR 437
              |   : || :| | | :  ||   :| |:|:|                     :| | |
Db        636 EELAEVEADIERIEKKLSQLK---QKYNEEEYKKKREEKEELEKELARLEAQKKELEKRR 692

Qy        438 DFINKQADAYVEKGILDKEEAEAIKR 463
              | |     : :||   :||  | :|:
Db        693 DTIK----STLEKLKAEKENRERVKK 714
```

```
RESULT 11
BAG_STRAG
ID   BAG_STRAG      STANDARD;      PRT;  1164 AA.
AC   P27951;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   IgA FC receptor precursor (Beta antigen) (B antigen).
GN   BAG.
OS   Streptococcus agalactiae.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1311;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
RC   STRAIN=LA239;
RX   MEDLINE=91312121; PubMed=1857207;
RA   Jerlstroem P.G., Chhatwal G.S., Timmis K.N.;
RT   "The IgA-binding beta antigen of the c protein complex of Group B
RT   streptococci: sequence determination of its gene and detection of two
RT   binding regions.";
RL   Mol. Microbiol. 5:843-849(1991).
RN   [2]
RP   IDENTIFICATION OF IG-LIKE DOMAIN.
RX   MEDLINE=97035265; PubMed=8880921;
RA   Bateman A., Eddy S.R., Chothia C.;
RT   "Members of the immunoglobulin superfamily in bacteria.";
RL   Protein Sci. 5:1939-1942(1996).
CC   -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC       an amide bond (Potential).
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59771; CAA42442.1; -.
DR   PIR; S15330; FCSOAG.
DR   InterPro; IPR004829; Csurface_antigen.
DR   InterPro; IPR005877; Gpos_YSIRK.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR006192; LPXTG.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF05062; RICH; 1.
DR   Pfam; PF04650; YSIRK_signal; 1.
DR   ProDom; PD153432; Csurface_antigen; 1.
DR   SMART; SM00409; IG; 1.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR   PROSITE; PS50835; IG_LIKE; FALSE_NEG.
KW   Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
KW   Immunoglobulin domain.
FT   SIGNAL        1     37
FT   CHAIN        38   1135       IGA FC RECEPTOR.
FT   PROPEP     1136   1164       REMOVED BY SORTASE (POTENTIAL).
FT   DOMAIN      434    534       IG-LIKE.
FT   DOMAIN      199    438       IGA-BINDING (POTENTIAL).
FT   DOMAIN      439    826       IGA-BINDING (POTENTIAL).
FT   DOMAIN      827    945       PRO-RICH REPEATS.
FT   SITE       1132   1136       LPXTG SORTING SIGNAL (POTENTIAL).
FT   MOD_RES    1135   1135       AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ   SEQUENCE   1164 AA;  131051 MW;  65DE94AF720A5474 CRC64;

  Query Match             7.2%;  Score 172.5;  DB 1;  Length 1164;
  Best Local Similarity   21.6%;  Pred. No. 0.3;
  Matches  124;  Conservative   77;  Mismatches  189;  Indels  185;  Gaps   25;

Qy         36 LSAERPLNE--------QIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLLKAIT-EKEK 86
              |  |   ||         |  | |:||  |     : |  : : : ||  |: | :  :| :
Db        149 LELENQFNETNRLLHIKQHEEVEKDKKAK----QQKTLKQSDTKVDLSNIDKELNHQKSQ 204

Qy         87 IEK--ERQSIRSSPLDNKL-------------NVEDVDSTKNRKLIDDYDSTKSGLDHK 130
              :||  |:: | :   |: |                || :     :|   : |||:||| :
Db        205 VEKMAEQKGITNEDKDSMLKKIEDIRKQAQQADKKEDAEVKVREELGKLFSSTKAGLDQE 264

Qy        131 FQDDPDGLHQLDGTPLTAEDIVHKI-----------AARIYEENDRAVFDKIVSKLLNLG 179
               |:      |    |  ::|:   |:            | :  || |:|  ::  ::: |
Db        265 IQE-----HVKKET--SSEENTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQF 317

Qy        180 L-------------------ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEE 210
              |                   : :: : :||     |:        |: |::|        ||
Db        318 LENAQKLKEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEG 377

Qy        211 DPNKPTSWTE-------------------NQAGKIPEKVTPMAAIQDGLAKGENDETVSN 251
                 |: :  |                   :  |  |:|||   ::  ||  ||::
Db        378 KLNQSKNLPELKQLEEEAHSKLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQ---Q 434

Qy        252 TLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVK 311
               : ||   |  |    ||          |    || ||               || :||
Db        435 KIELTVSPENITVYEGED--------VKFTVTAKS-DS-------------KTTLDFSD 471

Qy        312 MMVKYG-TISPEEGVSYLENLD-----------------EMIALQTK----NKLEKNATD 349
              :: ||  ::|     :|  | |                 : : |: |     | :|| |
Db        472 LLTKYNPSVSDRISTNYKTNTDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTI 531

Qy        350 NISKLFPAPSEKSHEETDSTKEEAA----------------------KMEKEYGSLKD 385
               : |        |: |: ||  ||                      |:||     | :
Db        532 TVQKKEEKQVPKTPEQKDSKTEEKVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELME 591

Qy        386 STKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKK----HDKKGNKEDYDLSKMRDFIN 441
               :  |||  :    |   |:  | |:: |   ||       |   | |       ||:|
Db        592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMN 648

Qy        442 KQADAYVEKGILDKEEAE---------AIKRIYSS 467
               |  | :|  :| ::  :          ||:|: |
Db        649 YQLHAQME--MLTRKVVQYMNKYPDNAEIKKIFES 681
```

```
RESULT 12
TRDN_CANFA
ID   TRDN_CANFA     STANDARD;      PRT;   700 AA.
AC   P82179;
DT   28-FEB-2003 (Rel. 41, Created)
```

```
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Triadin.
GN   TRDN.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Heart, and Skeletal muscle;
RX   MEDLINE=99428545; PubMed=10497235;
RA   Kobayashi Y.M., Jones L.R.;
RT   "Identification of triadin 1 as the predominant triadin isoform
RT   expressed in mammalian myocardium.";
RL   J. Biol. Chem. 274:28660-28668(1999).
CC   -!- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRIN TO THE
CC       JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC       COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC       reticulum.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=Skeletal;
CC         IsoId=P82179-1; Sequence=Displayed;
CC       Name=Cardiac 1;
CC         IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
CC       Name=Cardiac 3;
CC         IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
CC   -!- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF165916; AAF00222.1; -.
DR   EMBL; AF165915; AAF00221.1; -.
DR   EMBL; AF165917; AAF00223.1; -.
KW   Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW   Alternative splicing.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     47     67       POTENTIAL.
FT   DOMAIN       68    700       LUMENAL (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    257    277       DQYAFCRYMIDMFVHGDLRPG -> GKHSEEVAGGSKRTLG
FT                                KKQIQ (in isoform Cardiac 1).
FT                                /FTId=VSP_004001.
FT   VARSPLIC    278    700       Missing (in isoform Cardiac 1).
FT                                /FTId=VSP_004002.
FT   VARSPLIC    466    466       E -> EPIKGKEVKVPGSLKEKE (in isoform
FT                                Cardiac 3).
FT                                /FTId=VSP_004003.
FT   VARSPLIC    524    579       EEKVVKQVKATEKAAIEKTVKPKPAKKAEHQEKESPTIKTD
FT                                KPKPTSKETPEVTES -> GILQVVPVVLNCLFLVQFQQDE
FT                                ELNVESKVFRMIHVLSHPTSRTSPILVISTTCRT (in
FT                                isoform Cardiac 3).
FT                                /FTId=VSP_004004.
SQ   SEQUENCE   700 AA;  78152 MW;  F033E3AA1BEE0C56 CRC64;

  Query Match             7.0%;  Score 169;  DB 1;  Length 700;
  Best Local Similarity  20.0%;  Pred. No. 0.24;
  Matches   97;  Conservative   75;  Mismatches  176;  Indels  136;  Gaps   19;

Qy         27 QDKSLH----NRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT 82
              : | :|     :|   || :  ::|  |::|:|:    | |                 |
Db        139 KQKEIHKEKAEKEEKPERKILAKVAHREKEKVKEKEKSEKK----------------AT 181

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLD 142
              ||||||: :       : |   ::    |  : |      | |    |  :    : ::
Db        182 HKEKIEKKEKP------ETKTMAKEERKAKTEEKIK--KEVKGGKQEKVKPTAAKVKEVQ 233

Qy        143 GTPLTAEDIVHKIAARI--YEENDRAVFDKIVSKLLNLGLITESQAHTL---------- 189
              ||  |::   |  | :  :|: |:  | : :  :   | :   |:  |
Db        234 KTPPKAKEKEGKETAAVAKHEQKDQYAFCRYMIDMFVHGDLRPGQSPALPPPLPTVQASR 293

Qy        190 ----------EDEVAEVLQKLISKEANNYEEDPNKPTS-------WTENQAGKIPE--- 228
                       |:|  :  :| :: |    |::  |  |          : : || ||
Db        294 PTPASPTLEGKEEEEKKKAEKKVTSETKKKEKEDVKKKSDKDTAIDVEKKEPGKAPETKQ 353

Qy        229 ---KVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLK 285
                 ||  || :    | | |    | |    :|  |   ::  |                |
Db        354 GTIKVVAQAAAK----KDEKKEDSKKTKT---PVEEHPKGKKQEKKE------------K 394

Qy        286 SIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEK 345
              ::  | :|:: :                      :| |          : :  :|:   |:
Db        395 YVEPAKSSKKEHS--------------------APSE---------KQVKAKTERAKEE 424

Qy        346 NATDNISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDE--PKG 403
              :  :  | |    |  |:|  | |:  : ||   | | ||  :   |  |  |  ||  ||
Db        425 TSAASTKKAVPG---KKEEKTTKTVEQEIRKEK---SGKTSTASKDKEPEIKKDEKMPKA 478

Qy        404 KTEAYLEAIRKNIEWLKKHD---KKGNKEDYDLSKMRDFIN---------KQADAYVEKG 451
                |   :  :  ::  :| :   ||  | : |::|    ::          ||  |  ||
Db        479 DKEVKPKPPQSQVKKEEKSESQVKKEAKPEQDIAKPEKTVSHGKPEEKVVKQVKA-TEKA 537

Qy        452 ILDK 455
               ::|
Db        538 AIEK 541


RESULT 13
KF5C_HUMAN
ID   KF5C_HUMAN     STANDARD;      PRT;   957 AA.
AC   O60282; O95079;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
DE   2).
GN   KIF5C OR NKHC2 OR KIAA0531.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [2]
RP   SEQUENCE OF 355-585 FROM N.A.
RA   Engelender S., Sharp A.H., Colomer V., Tokito M.K., Lanahan A.,
RA   Worley P., Holzbaur E.L.F., Ross C.A.;
RT   "Huntingtin associated protein 1 (HAP1) interacts with the p150Glued
RT   subunit of dynactin.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, PROSTATE, AND
CC       TESTIS, AND MODERATE EXPRESSION IN KIDNEY, SMALL INTESTINE, AND
CC       OVARY.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
```

```
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB011103; BAA25457.1; -.
DR   EMBL; AF010146; AAD01436.1; -.
DR   HSSP; P56536; 2KIN.
DR   Genew; HGNC:6325; KIF5C.
DR   MIM; 604593; -.
DR   GO; GO:0005871; C:kinesin complex; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN       1    386       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN     406    923       COILED COIL.
FT   DOMAIN     859    956       GLOBULAR.
FT   DOMAIN     174    315       MICROTUBULE-BINDING.
FT   NP_BIND     86     93       ATP (BY SIMILARITY).
FT   CONFLICT   355    360       TLKNVI -> STHASV (IN REF. 2).
FT   CONFLICT   583    585       EFT -> DRV (IN REF. 2).
SQ   SEQUENCE   957 AA;  109494 MW;  A9F25BB1C994322A CRC64;

  Query Match             7.0%;  Score 167.5;  DB 1;  Length 957;
  Best Local Similarity  20.3%;  Pred. No. 0.42;
  Matches  103;  Conservative   98;  Mismatches  182;  Indels  125;  Gaps   23;

Qy         25 GSQDKSLHNR-----ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLL- 78
              | : |:: |       ||:||             : || |  | :  ::  : : :| :
Db        321 GQRAKTIKNTVSVNLELTAE------------EWKKKYEKEKEKNKTLKNVIQHLEMEL 367

Qy         79 ------KAITEKEKIEKERQSIRSSPLDNKLNVEDVD------STKNRKLIDDYDSTKSG 126
                  :|: | |:|  : |     | ||   ::::       ||: :   : ||   |
Db        368 NRWRNGEAVPEDEQISAKDQK-NLEPCDNTPIIDNIAPVVAGISTEEK---EKYDEEISS 423

Qy        127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEND-----RAVFDKIVSKLLNLGLI 181
              |  :  |  | ::|          :  |:  :: ::::     |  ::||  :|  | :
Db        424 LYRQLDDKDDEINQ-------QSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIE 476

Qy        182 TESQAHTLEDEVAEVLQKLISKEANNYE---EDPNKPTSWTENQAGKIPEKVTPMAAIQD 238
              |:     :||| |||| | : | ||:   ::     |    |     :: :| | :   |
Db        477 NEA----AKDEVKEVLQAL-EELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQR 531

Qy        239 GLAK-----------------------------GENDETVSNTLTLTNGLERR----TK 264
              |::                             | ||     ||   ||:       :
Db        532 ELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTND---VKTLADVNGVIEEEFTMAR 588

Qy        265 TYSEDNFEELQYFPNFYALLKS--IDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPE 322
              |      |::   |    |:|  :|| ::   |  :   : ||   :  :|  |
Db        589 LYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLT---- 644

Qy        323 EGVSYLENLDEMIALQTKNKLEK---NATDNISKLFPAPSEKSHEETDSTKE-------- 371
                |::|::     | : :||:   : :: ::||     || || :   ||
Db        645 ---DYMQNME-----QKRRQLEESQDSLSEELAKL--RAQEKMHEVSFQDKEKEHLTRLQ 694

Qy        372 EAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDY 431
              :| :|:|      :| :: :    :  :   : :  :: ||   : |:   :|    ||
Db        695 DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEK-LSSDY 753

Qy        432 DLSKMRDFINKQADAYVEKGIL--DKEE 457
                  :  |: |   :: :  :|| :|  || |  .
Db        754 NKLKIED---QEREMKLEKLLLLNDKRE 778


RESULT 14
MLP1_YEAST
ID   MLP1_YEAST     STANDARD;      PRT;  1875 AA.
AC   Q02455;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin-like protein MLP1.
GN   MLP1 OR YKR095W OR YKR415.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=93247549; PubMed=8483450;
RA   Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT   "A new yeast gene with a myosin-like heptad repeat structure.";
RL   Mol. Gen. Genet. 237:359-369(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC       REPAIR.
CC   -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC   -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L01992; AAA34783.1; -.
DR   EMBL; X73541; CAA51948.1; -.
DR   EMBL; Z28320; CAA82174.1; -.
DR   PIR; S38173; S38173.
DR   SGD; S0001803; MLP1.
DR   GO; GO:0005635; C:nuclear membrane; IDA.
DR   GO; GO:0005654; C:nucleoplasm; IDA.
DR   GO; GO:0006606; P:protein-nucleus import; IDA.
KW   Coiled coil; DNA repair.
FT   DOMAIN       69    487       COILED COIL (POTENTIAL).
FT   DOMAIN      531   1678       COILED COIL (POTENTIAL).
FT   DOMAIN     1834   1866       COILED COIL (POTENTIAL).
FT   CONFLICT    301    301       R -> A (IN REF. 1).
SQ   SEQUENCE   1875 AA;  218455 MW;  683A0D34C9066867 CRC64;

  Query Match             6.9%;  Score 166.5;  DB 1;  Length 1875;
  Best Local Similarity  23.6%;  Pred. No. 1;
  Matches  112;  Conservative   85;  Mismatches  178;  Indels   99;  Gaps   28;

Qy         30 SLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE-KEKIE 88
              :| : ::   |  : |   | ||  :| : | |:    ::|| |     :| :|| |  :|
Db        588 TLKSEKMDLESRIEELQKELEE--LKTSVPNED----ASYSNV----TIKQLTETKRDLE 637
```

```
Qy         89 KERQSIRSSPLDNKLNVEDVDSTKNRKLI-----DDYDSTKSGLDHKFQDDPDGLHQLDG 143
Db        638 SQVQD-----LQTRISQITRESTENMSLLNKEIQDLYDS-KSDISIK-------LGKEKS 684

Qy        144 TPLTAED---IVHKIAARIYEEND--RAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQ 198
Db        685 SRILAEERFKLLSNTLDLTKAENDQLRKRFDYLQNTILK----QDSKTHETLNEYVSCKS 740

Qy        199 KL--ISKEANNYEEDPNKPTSWTEN---QAGKI-PEK------VTPMAAIQ----DGLAK 242
Db        741 KLSIVETELLNLKEEQKLRVHLEKNLKQELNKLSPEKDSLRIMVTQLQTLQKEREDLLEE 800

Qy        243 ---------GENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEA 293
Db        801 TRKSCQKKIDELEDALSELKKETSQKDHHIKQLEEDNNSNIEWYQNKIEALK-------- 852

Qy        294 KEKETLIT-IMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNIS 352
Db        853 KDYESVITSVDSKQTDIEKLQYKVKSLEKE-----IE--EDKIRLHTYNVMDETINDDSL 905

Qy        353 KLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAI 412
Db        906 R-------KELEKSKINLTDAYSQIKEYKDLYETTSQSLQQTNSKLDE---SFKDFTNQI 955

Qy        413 RKNIEWLKKHDKKGNKED-YDLSKMRDF-INKQADAYVEKGILDKEEAEAIKRI 464
Db        956 -KNLT-----DEKTSLEDKISLLKEQMFNLNNELD--LQKKGMEKEKADFKKRI 1001


RESULT 15
CENE_HUMAN
ID   CENE_HUMAN     STANDARD;      PRT;  2663 AA.
AC   Q02224;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Centromeric protein E (CENP-E protein).
GN   CENPE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93024922; PubMed=1406971;
RA   Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT   "CENP-E is a putative kinetochore motor that accumulates just before
RT   mitosis.";
RL   Nature 359:536-539(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95196755; PubMed=7889940;
RA   Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT   "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT   microtubule motor.";
RL   EMBO J. 14:918-926(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
CC   -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC       KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC       OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC       AND/OR SPINDLE ELONGATION.
CC   -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC       CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC       QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z15005; CAA78727.1; -.
DR   PIR; S28261; S28261.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:1856; CENPE.
DR   GK; Q02224; -.
DR   MIM; 117143; -.
DR   GO; GO:0005699; C:kinetochore; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0008350; F:kinetochore motor activity; TAS.
DR   GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR   GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR   GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW   Cell cycle; Centromere.
FT   DOMAIN        1    335       KINESIN-MOTOR.
FT   DOMAIN      336   2471       COILED COIL (POTENTIAL).
FT   DOMAIN     2472   2663       GLOBULAR (POTENTIAL).
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
SQ   SEQUENCE   2663 AA;  312087 MW;  CEFC13880C8C8CB8 CRC64;

  Query Match             6.9%;  Score 166;  DB 1;  Length 2663;
  Best Local Similarity   21.8%;  Pred. No. 1.7;
  Matches  119;  Conservative  103;  Mismatches  175;  Indels  150;  Gaps   28;

Qy         26 SQDKSLHNRELSAERPLN---EQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT 82
Db       1102 AQEK---NHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQL-------LNVQEEMS 1151

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTK---NRKLIDDYDSTKS---------GLDHK 130
Db       1152 EMQKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVLKELQKS 1211

Qy        131 FQDDPDGL----HQLDGTPL-TAEDI----VH-KIAARIYEENDRAVFDKIVSKLLNLGL 180
Db       1212 FETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEK-TAQIINTQD 1270

Qy        181 ITESQAHTLEDEVAEVL---QKLIS--KEANNYEEDPNKPTSWTENQAGKIPEKVTPMAA 235
Db       1271 LEKS--HTKLQEEIPVLHEEQELLPNVKKVSETQETMNELELLTEQSTTK---DSTTLAR 1325

Qy        236 IQ----------------DGLAK-GENDETVSNTL-------------TLTNGLERRTK 264
Db       1326 IEMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSK 1385

Qy        265 TYSEDNFEEL-----------QYFPNFYALLK-------------SIDSEKE-AKEKE 297
Db       1386 QEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKD 1445

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIA--LQTKNKLE------KNATD 349
Db       1446 DLQRLQEVL-------------QSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEE 1491

Qy        350 NISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTE--- 406
Db       1492 TINELRVNLSEK--------ETEISTIQKQLEAINDKLQN-------KIQEIYEKEEQLN 1536

Qy        407 -AYLEAIRKNIEWLKKHDKKGNKEDYDL----SKMRDFINKQADAYVEKGILDKEEAEAI 461
Db       1537 IKQISEVQENVNELKQFKEHRKAKDSALQSIESKMLELTNRLQESQEEIQIMIKEK-EEM 1595
```

```
Qy        462 KRIYSSL 468
                ||:  :|
Db       1596 KRVQEAL 1602
```

Search completed: January 2, 2004, 12:34:53
Job time : 19 secs

OM protein - protein search, using sw model

Run on: January 2, 2004, 12:33:21 ; Search time 40 Seconds
(without alignments)
3019.215 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2062	86.0	466	11	Q8R1D7	Q8r1d7 mus musculu
2	1216	50.7	457	13	Q91847	Q91847 xenopus lae
3	198	8.3	7210	5	Q9V7G8	Q9v7g8 drosophila
4	198	8.3	9270	5	Q8MLD9	Q8mld9 drosophila
5	197	8.2	2081	10	Q9LH98	Q9lh98 arabidopsis
6	196.5	8.2	1175	16	Q8XNW6	Q8xnw6 clostridium
7	196	8.2	2139	5	Q07569	Q07569 entamoeba h
8	191.5	8.0	2760	5	Q8I5Y2	Q8i5y2 plasmodium
9	189	7.9	1434	5	Q8I492	Q8i492 plasmodium
10	188.5	7.9	1661	5	Q06166	Q06166 plasmodium
11	184	7.7	951	5	O96229	O96229 plasmodium
12	183.5	7.6	1790	3	Q07380	Q07380 saccharomyc
13	183.5	7.6	2612	5	Q8I5X5	Q8i5x5 plasmodium
14	180.5	7.5	3504	5	Q8IL45	Q8il45 plasmodium
15	179.5	7.5	1320	11	Q9JK25	Q9jk25 rattus norv
16	179.5	7.5	1510	5	Q25920	Q25920 plasmodium
17	179	7.5	853	10	Q9LFE4	Q9lfe4 arabidopsis
18	179	7.5	1558	5	O96275	O96275 plasmodium
19	179	7.5	1786	5	Q9U0P0	Q9u0p0 plasmodium
20	179	7.5	2269	5	Q26223	Q26223 plasmodium
21	179	7.5	2747	5	Q9BJX9	Q9bjx9 plasmodium
22	178.5	7.4	1323	5	Q9NB35	Q9nb35 plasmodium
23	178.5	7.4	3317	16	Q8EWP8	Q8ewp8 mycoplasma
24	178	7.4	840	5	Q8I5X4	Q8i5x4 plasmodium
25	177.5	7.4	1464	5	Q8IIF6	Q8iif6 plasmodium
26	177	7.4	1112	16	Q8EWQ1	Q8ewq1 mycoplasma
27	177	7.4	1871	10	Q9SRD5	Q9srd5 arabidopsis
28	176.5	7.4	682	5	Q8IBV6	Q8ibv6 plasmodium
29	175.5	7.3	1964	5	Q8SWQ7	Q8swq7 loligo peal
30	175	7.3	10578	5	Q8ISF5	Q8isf5 caenorhabdi
31	175	7.3	18519	5	Q8ISF6	Q8isf6 caenorhabdi
32	175	7.3	18534	5	Q8ISF7	Q8isf7 caenorhabdi
33	174.5	7.3	924	5	O15738	O15738 dictyosteli
34	174.5	7.3	1263	5	Q8IK49	Q8ik49 plasmodium
35	173.5	7.2	1795	5	Q8IE35	Q8ie35 plasmodium
36	172.5	7.2	1134	2	Q99051	Q99051 streptococc
37	172.5	7.2	1979	5	O96133	O96133 plasmodium
38	172	7.2	495	5	Q8IEK6	Q8iek6 plasmodium
39	171.5	7.1	2083	5	Q9N435	Q9n435 caenorhabdi
40	171.5	7.1	3484	5	P91257	P91257 caenorhabdi
41	171.5	7.1	5507	5	Q8IHN3	Q8ihn3 plasmodium
42	170.5	7.1	2771	5	Q26216	Q26216 plasmodium
43	170	7.1	1140	4	Q9ULE5	Q9ule5 homo sapien
44	170	7.1	1804	11	Q9Z1Q1	Q9z1q1 mesocricetu
45	170	7.1	3130	5	Q9BK46	Q9bk46 plasmodium

ALIGNMENTS

RESULT 1
Q8R1D7
ID Q8R1D7 PRELIMINARY; PRT; 466 AA.
AC Q8R1D7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to secretogranin III.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC024785; AAH24785.1; -.
SQ SEQUENCE 466 AA; 52783 MW; AB1CB96672763l5C CRC64;

Query Match 86.0%; Score 2062; DB 11; Length 466;
Best Local Similarity 87.0%; Pred. No. 2.1e-94;
Matches 410; Conservative 17; Mismatches 36; Indels 8; Gaps 2;

```
Qy          1 MGFLGTGTWILVLVL---PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKT 57
              |||| ||:|||||||   |||||||| ||||||||||||||||||||||||| |||||
Db          1 MGFLWTGSWILVLVLNSGPIQAFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKA 60

Qy         58 YPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI 117
                :| |:|| :|||| ||||||:|||||| :|||||||| | ||:|||| |||||||||
Db         61 FPSESKPSESNYSSVDNLNLLRAITEKETVEKERQSIRSPPFDNQLNVEDADSTKNRKLI 120

Qy        118 DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLN 177
              |:|||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db        121 DEYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLN 180

Qy        178 LGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
              |||||||||||||||| |||||||||||||  :|||| |||| |||||||||:||:|
Db        181 LGLITESQAHTLEDEVAEALQKLISKEANNYEETLDKPTSRTENQDGKIPEKVTPVAAVQ 240
```

```
Qy        238 DGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              ||    |||||||||||||:|||||||  : ||:|||||||||||||| ||||||||||||
Db        241 DGFTNRENDETVSNTLTLSNGLERRTNPHREDDFEELQYFPNFYALLTSIDSEKEAKEKE 300

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              |||||||||||||||||||||||||||||||||||||||| |||||||||||| ||:
Db        301 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDETIALQTKNKLEKNTTDS-----KT 355

Qy        358 PSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIE 417
              | ||| |||||||||||||||||||||||||||||||||| ||||||  ||||||||||||||
Db        356 PPEKSQEETDSTKEEAAKMEKEYGSLKDSTKDDNSNLGGKTDEATGKTEAYLEAIRKNIE 415

Qy        418 WLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||:|||||||||||||||||:|||||||||||||||| |||||||||
Db        416 WLKKHNKKGNKEDYDLSKMRDFINQQADAYVEKGILDKEEANAIKRIYSSL 466


RESULT 2
Q91847
ID   Q91847      PRELIMINARY;      PRT;   457 AA.
AC   Q91847;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Secretogranin III.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Int. pituitary;
RX   MEDLINE=96217355; PubMed=8632145;
RA   Holthuis J.C.M., Martens G.J.M.;
RT   "The neuroendocrine proteins secretogranin II and III are regionally
RT   conserved and coordinately expressed with proopimelanocortin in
RT   Xenopus intermediate pituitary.";
RL   J. Neurochem. 66:2248-2256(1996).
DR   EMBL; X92872; CAA63478.1; -.
SQ   SEQUENCE   457 AA;  51903 MW;  B3096E12D4B076D7 CRC64;

  Query Match           50.7%;  Score 1216;  DB 13;  Length 457;
  Best Local Similarity 57.0%;  Pred. No. 1.2e-52;
  Matches  269;  Conservative   63;  Mismatches  106;  Indels   34;  Gaps   10;

Qy          8 TWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQS 67
              | |: : : :| || |   ||| :|||||| |||| ||||||  | : :   ||:
Db          9 TAIIGVFVHVQGFPNPSIKQDKDVHNRELSEERPLEEQIAEA--DTVSRGTSTENQEMLK 66

Qy         68 NYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGL 127
              | || |:|  | :: ||:   |  | | |   |: ::| ||||| || : :|:|||
Db         67 NDSFADDLGFLMSMGEKDGDVKRSGSTRGSFGSNRGTLDDADSTKNLKLAEYFDTTKSMP 126

Qy        128 DHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH 187
              |:|::|||||||||  | |||||| ||| |||||||| ||||||||||||||||||:
Db        127 DYKYEDDPDGLHQLD-APFTAEDIVRKIATRIYEENDRGVFDKIVSKLLNLGLITESQAY 185

Qy        188 TLEDEVAEVLQKLISKEANNYEE--------DPNKPTSWTENQAGKIPEKVTPMAAIQDG 239
              |||||||||||:||: || | |:        |       : : ||:          :
Db        186 TLEDEVAEVLQQLIANEAKNREKEAGDLDYSDVRSDVYDDDKREGKM----------ET 234

Qy        240 LAKGENDET--VSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              | | |: |:    |  :|:  :||| :   ||   :|||||| |:||::|| : |||:
Db        235 LNKNEDRESSETKNEDSLST-VERRNELSPED---DLQYFPNFNRLVKSLNSENDVKEKK 290

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              ||:|||||||||||||||:|||||:||||||:|||:| |||| ||   : : |
Db        291 TLLTIMKTLIDFVKMMVKYGTITPEEGVNYLENLDDMIAVQVKNKLGKNFGSHRNAL--- 347

Qy        358 PSE-KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNI 416
              ||| |: :|:|||||||||||||  |||||| : ::      :|||| | |||||||||
```

```
Db        348 PSEDKNADESDSTKEEAAKMEKEYELLKDSTKPEPND--AAESKPKGKAETYLEAIRKNI 405

Qy        417 EWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              ||||:|:|: || | ||||:: |:|:|||:|||||||||||  ||||||||
Db        406 EWLKEHNKEANKGDNDLSKLKQFVNQQADSYVEKGILDKEEANVIKRIYSSL 457


RESULT 3
Q9V7G8
ID   Q9V7G8      PRELIMINARY;      PRT;  7210 AA.
AC   Q9V7G8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG18255 protein.
GN   STRN-MLCK OR CG8304 OR CG18255.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
```

```
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAF58087.2; -.
DR   FlyBase; FBgn0013988; Strn-Mlck.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 13.
DR   SMART; SM00409; IG; 21.
DR   SMART; SM00408; IGc2; 20.
DR   PROSITE; PS50835; IG_LIKE; 20.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   7210 AA;  808965 MW;  6B2E7395C07140D2 CRC64;

  Query Match             8.3%;  Score 198;  DB 5;  Length 7210;
  Best Local Similarity   23.8%;  Pred. No. 0.55;
  Matches  110;  Conservative   83;  Mismatches  180;  Indels   90;  Gaps   19;

Qy         35 ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSI 94
              |: :|:  :|:  |::: ::| :   | || ::          | |   |:||:| :::
Db       4757 EIVSEKITDEKAQESQKKEVKGS---EAKPKKAKV-------LEKKSIEEEKLEDKKEKQ 4806

Qy         95 RSSPLDNKLNVEDVDSTKNRKLIDD--YDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIV 152
                | :| |    :|    : |: |:   :| |  :    :  |      |:   :  | :
Db       4807 TESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDS-EAKPKKAKVLEKKSIEEEKLE 4865

Qy        153 HK----IAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQ-KLISKEANN 207
              :|    : | |:: :|   :|||: :      |||   ::|  |:  : |:: |::
Db       4866 NKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIE 4925

Qy        208 YEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYS 267
              |:  :|    ||             :|| :   | |  |||| :|     : : :
Db       4926 EEKLEDKKEKQTE------------SAIDEKFQKAEVSETVSEKIT-----DEKAEESR 4967

Qy        268 EDNFEELQYFPNFYALL--KSIDSEK--EAKEKET------------LITIMKTLID-- 308
              ::  :: :  |    :|  |||: ||  : |||:|              |: : : |
Db       4968 KEEVKDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESAIDEKSQKAEVSETVSEKITDEK 5027

Qy        309 ---FVKMMVKYGTISP------EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPS 359
                 |  ||     |      |:    :| |||    ||: |:   |||  |:
Db       5028 AQESQKKEVKDSEAKPKKAKILEKKSIEIEKLDEKKEKQTETKV---ATDTKSQTVEVSE 5084

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWL 419
                 : ::   ||: |:|     ||||              | | |    ||      | |
Db       5085 IVLEKISEEKAEESQKVE-----LKDS-------------EAKSKKAKVLEKKSTLKEKL 5126

Qy        420 KKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIK 462
```

```
                ::|||  |||   :| :    : ||   ||  : :|:   ||
Db       5127 DENDKK-QKEDGATNKSQK--AEAADVVPEK--ISEEKVAEIK 5164


RESULT 4
Q8MLD9
ID   Q8MLD9      PRELIMINARY;      PRT;  9270 AA.
AC   Q8MLD9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG18255-PA.
GN   STRN-MLCK OR CG8304 OR CG18255.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
```

```
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAM70936.1; -.
DR   FlyBase; FBgn0013988; Strn-Mlck.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 29.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00408; IGc2; 29.
DR   SMART; SM00406; IGv; 4.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 26.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Immunoglobulin domain; Transferase.
SQ   SEQUENCE   9270 AA;  1036749 MW;  0DD82D040B27C2AD CRC64;

  Query Match             8.3%;  Score 198;  DB 5;  Length 9270;
  Best Local Similarity   23.8%;  Pred. No. 0.73;
  Matches  110;  Conservative   83;  Mismatches  180;  Indels   90;  Gaps   19;

Qy         35 ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSI 94
              |: :|:  :|:  |::: ::| :   | || ::         | |   |:||:| :::
Db       4757 EIVSEKITDEKAQESQKKEVKGS---EAKPKKAKV-------LEKKSIEEEKLEDKKEKQ 4806

Qy         95 RSSPLDNKLNVEDVDSTKNRKLIDD--YDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIV 152
                | :| |    :|    : |: |:   :| | :    :   |      |:   :  | :
Db       4807 TESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDS-EAKPKKAKVLEKKSIEEEKLE 4865

Qy        153 HK----IAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQ-KLISKEANN 207
              :|      : | |:: :|   :|||: :       |||   ::|  |:  : |:: |::
Db       4866 NKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIE 4925

Qy        208 YEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYS 267
               |:  :|    ||             :|| :   | |  ||||  :|     : : :
Db       4926 EEKLEDKKEKQTE------------SAIDEKFQKAEVSETVSEKIT-----DEKAEESR 4967

Qy        268 EDNFEELQYFPNFYALL--KSIDSEK--EAKEKET------------LITIMKTLID-- 308
              ::  :: :  |    :|  |||: ||  : |||:|             |: : : |
Db       4968 KEEVKDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESAIDEKSQKAEVSETVSEKITDEK 5027

Qy        309 ---FVKMMVKYGTISP------EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPS 359
                  |  ||      |       |:    :| |||    ||: |:   |||  |:
Db       5028 AQESQKKEVKDSEAKPKKAKILEKKSIEIEKLDEKKEKQTETKV---ATDTKSQTVEVSE 5084

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWL 419
                : ::   ||: |:|     ||||                | | |   ||      | |
Db       5085 IVLEKISEEKAEESQKVE-----LKDS------------EAKSKKAKVLEKKSTLKEKL 5126

Qy        420 KKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIK 462
               ::|||  |||   :| :    : ||  ||  : :|:   ||
Db       5127 DENDKK-QKEDGATNKSQK--AEAADVVPEK--ISEEKVAEIK 5164


RESULT 5
Q9LH98
ID   Q9LH98      PRELIMINARY;      PRT;  2081 AA.
AC   Q9LH98;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Genomic DNA, chromosome 3, BAC clone: T19N8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP002057; BAB03174.1; -.
SQ   SEQUENCE   2081 AA;  232851 MW;  D3603E1F85EFFF29 CRC64;

  Query Match             8.2%;  Score 197;  DB 10;  Length 2081;
  Best Local Similarity   25.2%;  Pred. No. 0.14;
  Matches  120;  Conservative   68;  Mismatches  175;  Indels  114;  Gaps   21;

Qy         27 QDKSLHNRELSAERPLNEQIAEAEED----------------KIKKTYPPENKPGQSNYS 70
              :|| | |:|   :   :: : : :|:                 | ||    |||  ::| :
Db       707 EDKKLENKESQTDSKDDKSVDDKQEEAQIYGGESKDDKSVEAKGKKKESKENKKTKTNEN 766

Qy         71 FVDNL--NLLKAITEKEKIEK--ERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSG 126
                | |   |:      | ||:||  :::|  :  :: | | : : ||:||    |    :||
Db       767 RVRNKEENVQGNKKESEKVEKGEKKESKDAKSVETKDN-KKLSSTENR----DEAKERSG 821

Qy        127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEN---DRAVFDKIVSKLLNLGLITE 183
               |:|   :    :|                : |:   ||   |  | :|  || |      |
Db       822 EDNKEDKEESKDYQ------------SVEAKEKNENGGVDTNVGNKEDSKDLKDDRSVE 868

Qy        184 SQAHTLED-----EVAEVLQKLISKE----ANNYEEDPNKPTSWTENQAGKIPEKVTPMA 234
               :|:  |       |  :   |  :||    ||| : |  | :   |:   |  ||
Db       869 VKANKEESMKKKREEVQRNDKSSTKEVRDFANNMDIDVQKGSG--ESVKYKKDEK----- 921

Qy        235 AIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAK 294
               ::|  | || :|: || :   | ::: |      |               |   :||
Db       922 --KEG-NKEENKDTI-NTSSKQKGKDKKKKKKESKN--------------SNMKKKEED 962

Qy        295 EKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKL 354
               :|| :   :|   |  |   |       :|     || |         || :| : |: ||
Db       963 KKEYVNNELKKQEDNKKETTKSENSKLKE-----ENKD--------NKEKKESEDSASK- 1008

Qy        355 FPAPSEKSHEETDS-TKEEAAKMEKEYGSLKDSTKDDNSNPG-------------GKTDE 400
                 :| :||  | |||||  | :|:    |   ||                     | :|
```

```
Db       1009 --NREKKEYEEKKSKTKEEAKKEKKKSQDKKREEKDSEERKSKKEKEESRDLKAKKKEEE 1066

Qy        401 PKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEE 457
Db       1067 TKEKKES------ENHKSKKKEDKKEHEDNKSMKKEED--KKEKKKHEESKSRKKEE 1115


RESULT 6
Q8XNW6
ID   Q8XNW6      PRELIMINARY;      PRT;  1175 AA.
AC   Q8XNW6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Probable exonuclease.
GN   SBCC OR CPE0216.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003185; BAB79922.1; -.
DR   InterPro; IPR003439; ABC_transporter.
KW   Exonuclease; Complete proteome.
SQ   SEQUENCE   1175 AA;  136878 MW;  7C6D2366525019C1 CRC64;

  Query Match             8.2%;  Score 196.5;  DB 16;  Length 1175;
  Best Local Similarity   22.6%;  Pred. No. 0.078;
  Matches  116;  Conservative   83;  Mismatches  196;  Indels  119;  Gaps   21;

Qy         29 KSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL-NLLKAITEKE-- 85
Db        256 KELYDKRIEEE----SLVSRSEEIKSFKERVEISNKADKVIVFINNLEEILKEINKEDLK 311

Qy         86 --KIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDG 143
Db        312 FSELNKKLEELINLREENKLKFEEFTKKKEEKLPDLRLKKEKLLES--QKERDILFQIKA 369

Qy        144 TPLTAEDIVHKI---------AARIYEENDRAVFDKIVSK---------------LLNLG 179
Db        370 DGVKLKEACKKIFEDRSKCDTKLNSIEENEKRLNEELKEKEERKEELFVHEEFKNKINSG 429

Qy        180 LITESQAHTLEDEVAEVLQKLI--SKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
Db        430 LFILNSYESLDKQFNEIKSEEVELKKYIKNLTEDKEKSEK-------DLKVKVESLSKIR 482

Qy        238 DGLAK------------------GENDETVSNTLTLTNGLERRTKTYSEDNFEE-LQYF 277
Db        483 DKLESLLKETPGDSNSILEKQIKLGEYREKLNKYKEIKNSLEESLKT--KNNFEEKLKTF 540

Qy        278 PNFYALLKSIDSEKEAKEKETLITIMKT--LIDFVKMMVKYGTISPEEGVSYLE------ 329
Db        541 ENQKLLL-----EKEVRELKDYINKVKVEELAHKLRENLVEGECCPVCGSTHHELNKVEK 595

Qy        330 -NLDE----MIALQTKNKLEKNATDNISKL---FPAPSEK------SHEETDSTKEEAAK 375
Db        596 INLEESNEKTILLESKEEKLKELILEFSKIEATLEYENKKIEELNISIEEVGEVNEERLK 655

Qy        376 -MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLS 434
Db        656 FLEEEFNTLKDKIEEFNLKK--------------ENLEKDLEKLK--EEKNNLE----- 693

Qy        435 KMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
Db        694 ---NIFNKAEVILCEKIVREKEIASKIKELDKEL 724
```

```
RESULT 7
Q07569
ID   Q07569      PRELIMINARY;      PRT;  2139 AA.
AC   Q07569; O02504;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Myosin heavy chain.
GN   MHCA.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HMI:IMSS;
RX   MEDLINE=93295430; PubMed=8515774;
RA   Raymond-Denise A., Sansonetti P., Guillen N.;
RT   "Identification and characterization of a myosin heavy chain gene
RT   (mhcA) from the human parasitic pathogen Entamoeba histolytica.";
RL   Mol. Biochem. Parasitol. 59:123-131(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HMI:IMSS;
RA   Guillen N.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L03534; AAB48065.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00242; MYSc; 1.
SQ   SEQUENCE   2139 AA;  245225 MW;  C68307341DB51DD1 CRC64;

  Query Match             8.2%;  Score 196;  DB 5;  Length 2139;
  Best Local Similarity   21.0%;  Pred. No. 0.17;
  Matches  107;  Conservative  108;  Mismatches  191;  Indels  104;  Gaps   23;

Qy         20 FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPEN-------KPGQSNYSFV 72
Db        859 FEEEGKKKDKEIED--------LKKKLAE----EIKKREAAENALASATAKTGELEAKIQ 906

Qy         73 D----------------------NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVD- 109
Db        907 DLEDKISELESKLSAAELDKQELNLKIENLEEDKEELKETIDNLKGDLKDSKLKGEDLEV 966

Qy        110 ---------STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIY 160
Db        967 EITELNSQINTLN-ATVDKKDKTIAEMQESIDEKEDEITKLKGDIKLLEEEKDDL----- 1020

Qy        161 EENDR----AVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPT 216
Db       1021 -EQDRADVSATKDDIAKKLNKITI----ECEDAKDEIAKLEQELEDEE--NKNKDLTNEL 1073

Qy        217 SWTENQAGKIPEKVTPMAAI-------QDGLAKG-ENDETVSNTLTLTNG-LERRTKTYS 267
Db       1074 QQTQLKLGETEKSLAAQVAATKKASDERDTLSQNLENEKLTTKNLTKTKADLEKKISGLK 1133

Qy        268 EDNFEELQYFPNFYALLKSID-----SEKEAKEKETLITIMKTLIDFV-KMMVKYGTISP 321
Db       1134 QD-YEDLEDDKN------KIEGDLRNAQRKIKELDDEITKGADVSQYLQKQKEEY----- 1181

Qy        322 EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK----ME 377
Db       1182 ESQIAKMQEEKEAIGNDVKNK-EKTIKE--KELEIQSLQEKLDETEVEKEDAEKKKKEIE 1238

Qy        378 KEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL-SKM 436
Db       1239 KEMKALQEE-KENVESSKNSTEKDKKKLEDNLKDTQKKLDDMTADNEKLKAKAKDLEAQL 1297
```

```
Qy        437 RDFINKQADAYVEKGILDKEEAEAIKRIYS 466
              :  :    |  :  :|:|::|:: | : |
Db       1298 NEVQDNHEKAVADAELLNKKKAQSDKELNS 1327
```

RESULT 8
Q8I5Y2
ID Q8I5Y2 PRELIMINARY; PRT; 2760 AA.
AC Q8I5Y2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL0315C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014845; AAN36152.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2760 AA; 330896 MW; EE9964C4845181AC CRC64;

Query Match 8.0%; Score 191.5; DB 5; Length 2760;
Best Local Similarity 21.1%; Pred. No. 0.37;
Matches 121; Conservative 92; Mismatches 207; Indels 153; Gaps 24;

```
Qy         29 KSLHNRELSAERPLNEQIAEAEE--------------DKIKKTYPPENKPGQSNYSFVD 73
              |:    |   |: :|:|:   :|                ::|| |   | : : |
Db       1452 KNEAKEEEETEKNINDQMVNEKEMDVNSKNEREIVQVHNEIKNTNNKEEEGKKKNLLKEK 1511

Qy         74 NLN--LLKAITEKEKIEKERQ--SIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDH 129
              :|  |   | :::| ||::   ::   |: : |  :::   : | :   |     |:
Db       1512 EINDCLNDYINKQKKKEKKKNNWAMYGRPIVKRQNNRNINIKNDLKKLYSSKSESGFNDY 1571

Qy        130 KF------------QDDPDGLHQLDGTPLTAE---DIVH---------KIAARIYEEN-- 163
              |            :|| |  :|       |   ||:|          |:   ||| :
Db       1572 AFYAERFFEVITGYNSEPDYLSDIDNQAKNEENKNDIIHNNNIIKISKKMKENIYENSPF 1631

Qy        164 ---DRAVFD-------------------------KIVSKLLNLGLITESQA-------- 186
                | :::                          | ::| :::   |:  :
Db       1632 HTYGRPIYEKKSKNPNNYNKIKSTTHNAILKKKRKKTLNKSISINSFTKMNSSNNKIVKR 1691

Qy        187 -----HTLEDEVAEVLQKLISKEANNYEEDP----------------NKPTSWTENQAGK 225
                 :|:::     ::|:| || |   ||                  ::     |: :  |
Db       1692 TSIKNNTIDNYNNSTIKKIIHKEQN--VEDQGYIDLKTKRKLIYDALDEINQQTQQKNLK 1749

Qy        226 IPEKVTPMAAIQDGLAKGENDETVSNTLTLTNG--LERRTKTYSEDNFEELQYFPNFYAL 283
              | | :|     | |   |:|   ||| :  |    :|: |:    || || |:     ||
Db       1750 ITENIT-----QVGKKHGQN---VSNIIKNTGAMLIEKITRK-GNDNNEEDQFSEELKAL 1800

Qy        284 LKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEM--------I 335
              | :   || |  | | |: :   :  | : :  ::  ||      :|::||        |
Db       1801 EK-LKKLKELKRIEELKTLEE---EEKKRLEELNSLKVEEEKKKKKNMEEMKKMEEMKKI 1856

Qy        336 ALQTKNKLEK-----NATDNISKL--------FPAPSEKSHEETDSTKEEAAKM--EKEY 380
              || | : ||       : | ||              | :  |    :||  ||  |||
Db       1857 ELQRKKEEEKLLQYSRRQEQIRKLEELKKKEELKLIEEMNKLEEQRKREEIKKMEEEKER 1916
```

```
Qy        381 GSLKDSTKDDNSNPGGKTDEPKGKTEAYL-------EAIRKNIE--WLKKHDKKGNKEDY 431
              ||   :         | :| : | :  |       |  || :|  ||::  :|   :
Db       1917 EKLKKIKEQQKKEEMKKQEEQRKKEQIKLMEEMRKFEEQRKKVELKWLEEEKRKEELKKL 1976

Qy        432 DLSKMRDFINKQADAYVEKGILDKEEAEAIKRI 464
              :  | |:   |: :   :|   :|:  | ||::
Db       1977 EEQKKRE-ERKKMEEEEKKREEEKKREEEIKKM 2008
```

RESULT 9
Q8I492
ID Q8I492 PRELIMINARY; PRT; 1434 AA.
AC Q8I492;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mature parasite-infected erythrocyte surface antigen (MESA) or
DE PfEMP2.
GN MESA.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51374.1; -.
SQ SEQUENCE 1434 AA; 168287 MW; AB0005F9DC26C989 CRC64;

Query Match 7.9%; Score 189; DB 5; Length 1434;
Best Local Similarity 21.1%; Pred. No. 0.23;
Matches 101; Conservative 96; Mismatches 180; Indels 102; Gaps 23;

```
Qy         27 QDKSLHNRELSAERPLNEQ-----IAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI 81
              :|| |   :   :  |::    : | ::: :|     |   |: :  :  : | |
Db        501 KDKVLGEGDKEDVKEKNDEQKDKVLGEGDKEDVK-----EKNDGKKD-KVIGSEKTQKEI 554

Qy         82 TEKEKIEKE-----RQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQD--- 133
              |||:||      :: ::   :|  : |  |  |  ::|   :  |  :  : :|
Db        555 --KEKVEKRVKKKCKKKVKKGIKEN--DTEGNDKVKGPEII--IEEVKEEIKKQVEDGIK 608

Qy        134 --DPDGLHQLDGTPLTAEDIVHKIAARIYE---ENDRAVFDKIVSKLLNLGLITESQAHT 188
                | :|  :: |  :  |::  :|  :: |   |||    ||:        :|||
Db        609 ENDTEGNDKVKGPEIITEEVKEEIKKQVEEGIKENDTEGNDKVKGP----EIITEEVKEE 664

Qy        189 LEDEVAEVLQ--------KLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              :: :| | ::         |||:|     |         : |||:     :||     | : :
Db        665 IKKQVEEGIKENDTESKDKLIGQEIITEEVKEGIKENDTENK-----DKVIGQEIITEEV 719

Qy        241 AKG--ENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKET 298
              :|  |||              | : |   ::   |          : | |: ::|   ||
```

```
Db        720 KEGIKEND------------TENKDKVIGQEIITE--------EVKKEIEKQEEKGNKEN 759

Qy        299 LITI------MKTLIDFVKMMVKYGTISPEEGV--SYLENLD-----EMIALQTKNKLEK 345
              :: |       : :|: || ::|      |:|:  :: |: |     |:|  : | ::||
Db        760 ILEIKDIVIGQEVIIEEVKKVIKK---KVEKGIKENHTESKDKVIGQEIIVEEVKEEIEK 816

Qy        346 NATDNISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKT 405
                : |             |    :|::    |   |   |  ::   |     | : | |
Db        817 QVEEGIK----------ENDTESKDKVIGQEVIKG---DVNEEGPENKDKVTKQEKVK- 861

Qy        406 EAYLEAIRKNIEWLKKHDKKGNKEDYDLSK--MRDFINKQADAYVEKGILDKEEAEAIK 462
              |   |  :|  : :|| : |  ::|  : |  |:: :|::  |  :| |  ::| | :|
Db        862 EVKKEVKKKVKKRVKKRNNKNERKDNVIGKEIMKEDVNEKDTANKDKEIEQEKEKEEVK 920


RESULT 10
Q06166
ID   Q06166      PRELIMINARY;      PRT;  1661 AA.
AC   Q06166;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Mature PARASITE-infected erythrocyte surface antigen (Antigenic
DE   protein PFEMP2).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PALO ALTO;
RX   MEDLINE=92158014; PubMed=1741020;
RA   Coppel R.L.;
RT   "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT   binds human erythrocyte protein 4.1.";
RL   Mol. Biochem. Parasitol. 50:335-347(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PALO ALTO;
RA   Kun J.F.J., Waller K.L., Coppel R.L.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 797-850 FROM N.A.
RX   MEDLINE=93122844; PubMed=1478701;
RA   Saul A., Yeganeh F., Howard R.J.;
RT   "Conservation of repeating structures in the PfEMP2/MESA protein of
RT   Plasmodium falciparum.";
RL   Immunol. Cell Biol. 70:353-355(1992).
DR   EMBL; AF056936; AAC13303.1; -.
DR   EMBL; S52458; AAB24869.1; -.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Antigen.
SQ   SEQUENCE   1661 AA;  195479 MW;  AF340527D85A9D29 CRC64;

  Query Match             7.9%;  Score 188.5;  DB 5;  Length 1661;
  Best Local Similarity   19.9%;  Pred. No. 0.29;
  Matches   94;  Conservative   90;  Mismatches  189;  Indels   99;  Gaps   15;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI 81
              |   ::|| :   :   |    |:: | || | |:    : :                 :
Db        895 KDTANKDKEIEQEKEKEEVKEKEEVKEKEEVKEKEEVKEKEE----------------V 937

Qy         82 TEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQL 141
                |||:::: :      :  |  |:: |:    | |:  :  |  :    : | :   ::
Db        938 KEKEEVKEKEEVKEKEEVKEKEEVKEKDTESKDKEIEQ-EKEKEEVKEVKEKDTENKDKV 996

Qy        142 DGTPLTAEDIVHKIAARIYEEN------DRAVFDKIVSKLLNLGLITESQAHTLEDEVAE 195
                |  :  |:|  ::  |: : |      |  :  :|::::|     : :    | |  |
Db        997 IGQEIIIEEIKKEVKKRVKKRNNKNENKDNVIVQEIMNEMLRKDTANKDKVIEQEKEKEE 1056

Qy        196 VLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTL 255
              | :|   ||    :|                 |:|             | :|
Db       1057 VKEKEEVKEKEEVKE----------------KEEVK-----------EKEEV------- 1081

Qy        256 TNGLERRTKTYSEDNPEELQYFPNFYALLKSIDSEKEAKEKET----LITIMKTLIDFVK 311
                    :   | |:|   | :       |   :  || |||:|     :   : :|: :|
Db       1082 -----KEKDTESKDKEIEQE---------KEKEEVKEVKEKDTENKDKVIGQEIIIEEIK 1127

Qy        312 MMVKYGTISPEEGVSYLENLDEMIALQTKNK-LEKNATDNISKLFPAPSEKSH-EETDST 369
                ||      ::  :  || | :|  :  |: : :  | |  |:     ||   :| :
Db       1128 KEVKKRV---KKRNNKNENKDNVIVQEIMNEDVNEKDTANKDKVIEQEKEKEEVKEKEEV 1184

Qy        370 KEEAAKMEKEYGSLKDS-----------TKDDNSNPGGKTDEPKGKT---EAYLEAIRKN 415
                ||:    |||    ||:           |::      |  | | |    |  :| |:|
Db       1185 KEKEEVKEKEEVKEKDTESKDKEIEQEKEKEEVKEVKEKDTENKDKVIGQEIIIEEIKKE 1244

Qy        416 IE--WLKKHDKKGNKEDYDLSK-MRDFINKQADAYVEKGILDKEEAEAIKRI 464
              ::    |:::|  ||::  : : | : :|::  |  :| |  ::| | :| :
Db       1245 VKKRVKKRNNKNENKDNVIVQEIMNEDVNEKDTANKDKVIEQEKEKEEVKEV 1296


RESULT 11
O96229
ID   O96229      PRELIMINARY;      PRT;   951 AA.
AC   O96229;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFB0680W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=99021743; PubMed=9804551;
RA   Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA   Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA   Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA   Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA   Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE001410; AAC71925.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   951 AA;  112486 MW;  AC8D889358A84F4F CRC64;

  Query Match             7.7%;  Score 184;  DB 5;  Length 951;
  Best Local Similarity   22.0%;  Pred. No. 0.25;
  Matches  104;  Conservative   76;  Mismatches  182;  Indels  110;  Gaps   17;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEK 84
              | || |  | |    :: : | : | || |.::    ::|  ::  |           :
Db        161 GKQDISNSNAE--NKKDVKEGVKELEEKKKEEKISDDHKVEENKKS-----------DD 206

Qy         85 EKIEKERQSIRSSPLDNKLN----VEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQ 140
```

```
Db        207 HKVEENKKSDDHKVEENKKSDDHKIEEVKKVEEHEEDEEEDKKEKKSENKNKDENKDEND 266

Qy        141 LDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKL 200

Db        267 EDNDEISDEDEVDDDVEEDKNENDDIDDDK----------KETDKTHLEEEENEIIEKE 315

Qy        201 IS-KEANNYEEDPNKPTSW-TENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNG 258

Db        316 FSDKKKNGKNKDTKKEKSKDTEKEKSKDIEK----------EKSKDKEKEKS------- 357

Qy        259 LERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGT 318

Db        358 -KDKEKEKGKDKEKE----------KSKDIEKE-KEKDKDIEKEKSK-DTAKEKEKDKD 403

Qy        319 ISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSH----------E 364

Db        404 IEKEKS-------KDMEKLKNKQNDEKKKDDNEKK----KNDKQDIHDDNDDENDMEEIE 452

Qy        365 ETDSTKEEAAKME----------KEYGSLKDSTK-DDNSNPGGKTDEPKGKTEAYLEAI 412

Db        453 ENDDEEDEDEDMENKKKKKKKGKNGNENGNENGSENGNENGNENGNENENKNESENENENE 512

Qy        413 RKNIEWLKKHDKKGNKEDYDLSKMRDF----------INKQADAYVEKGILD 454

Db        513 NENENGNENENEKENKDKNIKEIENVTNANKENYEKINKNSEITITKSNID 564


RESULT 12
Q07380
ID   Q07380      PRELIMINARY;      PRT;  1790 AA.
AC   Q07380; P89892;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 206.5 kDa protein YDL058W.
GN   USO1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bloecker H., Brandt P.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z74106; CAA98621.1; -.
DR   EMBL; Z74105; CAA98620.1; -.
DR   SGD; S0002216; USO1.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR006955; Uso1_p115_C.
DR   InterPro; IPR006953; Uso1_p115_head.
DR   Pfam; PF04871; Uso1_p115_C; 1.
DR   Pfam; PF04869; Uso1_p115_head; 1.
SQ   SEQUENCE   1790 AA;  206450 MW;  90062544F55A52EE CRC64;

  Query Match             7.6%;  Score 183.5;  DB 3;  Length 1790;
  Best Local Similarity   20.3%;  Pred. No. 0.56;
  Matches  120;  Conservative   91;  Mismatches  200;  Indels  179;  Gaps   23;

Qy         29 KSLHN--RELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI----T 82

Db        965 KSLANNYKDMQAE---NESLIKAVEE----------SKNESSIQLSNLQNKIDSMSQ 1008

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI------DDYDSTKSGLDHKFQD--- 133

Db        1009 EKENFQIERGSIEKNIEQLKKTISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATT 1068

Qy        134 -DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH----- 187

Db        1069 ANDENVNKISELTKTREELEAELAA--YKNLKNELETKLETSEKALKEVKENEEHLKEEK 1126

Qy        188 -TLEDEVAEVLQKLISKEAN----------NYEED-PNKPTSWTENQAGKIPEK 229

Db        1127 IQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKERQYNE-EISQLNDE 1185

Qy        230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRT----------KTYSEDN-- 270

Db        1186 ITSTQQENESIKK-KNDELEGEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEA 1244

Qy        271 ----------FEELQYFPNF----------YALLKSIDSEKEAK 294

Db        1245 SLLESIKSVESETVKIKELQDECNFKEKEVSELEDKLKASEDKNSKYLELQK-ESEKIKE 1303

Qy        295 EKETLITIMKTLIDFVKMMVK---------- 315

Db        1304 ELDAKTTELKIQLEKITNLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQ 1363

Qy        316 ----------YGTISPE--EGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEK 361

Db        1364 AFEKERKLLNEGSSTITQEYSEKINTLE--DELIRLQNENELKAKEIDNTRSELEKVSLS 1421

Qy        362 SHEETDSTKEEAAKMEKEYGSLKDS-TKDDNSNPGGKTDEP------KGKTEAYLEAIRK 414

Db        1422 NDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAK 1481

Qy        415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILD-KEEAEAIKR 463

Db        1482 VEEGLKKLEEESSKEKAELEKSKEMM-KKLESTIESNETELKSSMETIRK 1530


RESULT 13
Q8I5X5
ID   Q8I5X5      PRELIMINARY;      PRT;  2612 AA.
AC   Q8I5X5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL0350C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014845; AAN36159.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2612 AA;  311421 MW;  C950B6E466E3D1C5 CRC64;

  Query Match             7.6%;  Score 183.5;  DB 5;  Length 2612;
  Best Local Similarity   22.3%;  Pred. No. 0.87;
  Matches  114;  Conservative   98;  Mismatches  196;  Indels  103;  Gaps   23;

Qy         27 QDKSLHNRELSAERP---LNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE 83

Db        991 REKEIDNERRNIERKENFLNHTKKELEEEKL----------NNMKEKNELEMLK--KE 1036

Qy         84 KEKIEKERQSIRSSPLDNKLNVE----------DVDSTKNR--KLIDDYDSTKSG 126

Db        1037 LESLEKEKKQIIECEYNNLQNKEEEIQRNERNNLIKQKELNSRMDRYNELLEDLNKKRKE 1096

Qy        127 LD---HKFQDDPDG-----LHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNL 178
```

```
Db       1097 LEMEKMKLLDDVQEERIKFLNEKNNMQKEKENEINYMKEELRKERILMIEEVEKMKVIML 1156

Qy        179 GLI----------TESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPE 228

Db       1157 EDIEKNKEKMIKNVEKENEKLKDEIEKERRNMI----QNLEEEKKEFKLYLEQKYKENFE 1212

Qy        229 KVTPMAAIQDGLAK--GENDETVSNTLTLTNGLERRTKTYSEDNFEELQ--YFPNF---- 280

Db       1213 NE------KSGLAKKFDEENEKLQNEI----GNEKRKLHKERDNFEEQKKIYEEEFRNKC 1262

Qy        281 ----YALLKSID--SEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYL------ 328

Db       1263 EKYEESIQKKYDVLEEDKNKMKYLIIKEQEELENYKKNI--YLDIEEEKEKLYVQQEKLN 1320

Qy        329 ---ENL---DEMIALQTKN-KLEKNATDNISKL----FPAPSEKSHEETDSTKEEAAKME 377

Db       1321 LEKENLLVEKEQIEIELKNYKNFKEKEENDIKIKIINLSQQQEDLNKEKENIEKEKEQIE 1380

Qy        378 KEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKK-GNKEDYDLSKM 436

Db       1381 KRKYDL-DEREESLNNDKAQMEESRKIFDEQLDKIKKNKEELLNYDKELKNREIHLMEKE 1439

Qy        437 RDFINKQADAYVEKGILDKEEAEAIKRIYSS 467

Db       1440 KEYNKKLEESNKKKESLNSFDVEL--KEYSS 1468
```

RESULT 14
Q8IL45
ID Q8IL45 PRELIMINARY; PRT; 3504 AA.
AC Q8IL45;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0404.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014822; AAN37017.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3504 AA; 408303 MW; B8454D48D55BE4F0 CRC64;

Query Match 7.5%; Score 180.5; DB 5; Length 3504;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 104; Conservative 93; Mismatches 172; Indels 139; Gaps 22;

```
Qy         32 HN----RELSAE-------RPLNEQIAEAE-----EDKIKKTYPPENKPGQ--------- 66

Db        982 HNLIGKNEVSDETMNEYIISPKSEIINEQENTPQYEEVLSKNINNDKDNGKRINNNDQLE 1041

Qy         67 ----SNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDN----KLNVEDVDSTKNRKLID 118

Db       1042 SPILSNEKKTDDIHI---IEEKEKIDK---PIHNNDLENQNHEKKHIQD-------DKLE 1088

Qy        119 DYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNL 178

Db       1089 EYSTLPNSKEMK-------------NISGDNIENNLLSNKKEENHKIIHENIEEDTTSK 1134

Qy        179 GLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTEN-----QAGKI------- 226

Db       1135 DNEKNMQNYIIHDDLG---KKDISRENTGEDISPNKYLKNKENYGILDDKGKIKIYEENN 1191

Qy        227 PEKVTPMAAIQDGLAKGENDETV----------------SNTLTLTNGLERRTKTYSEDN 270

Db       1192 EEYVTDVGSGENKVSEGNSNDNIMPEEYGKNKNSQENSEDNIMTEKNGKNKSSEENTDDN 1251

Qy        271 FEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLEN 330

Db       1252 ----------------IMTEKNGKDKSSEENANDNI-----MTEKNGKNKSSQG-----N 1285

Qy        331 LDEMIALQ---TKNKLEKNATDNISKLFPAPSEKSHEETDST-------KEEAAK----- 375

Db       1286 IDDNIMTEKNGKNNNSQENSDDNIMTEEYEKNKNSQENTDDNIMTEKNGKNKSSEENTDD 1345

Qy        376 --MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL 433

Db       1346 NIMTEKNG--KDKSSEENANDNVMT-EKNGKNKSSEENADDNIMTEKNGKNKSSEENADD 1402

Qy        434 SKMRD--FINKQADAYVEKGILDKEEAE 459

Db       1403 NIIREENGKNKSSQENSDDNIMTEEYGE 1430
```

RESULT 15
Q9JK25
ID Q9JK25 PRELIMINARY; PRT; 1320 AA.
AC Q9JK25;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CLIP-170.
GN CYLN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RA Galjart N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RX MEDLINE=21185938; PubMed=11290329;
RA Akhmanova A., Hoogenraad C.C., Drabek K., Stepanova T., Dortland B.,
RA Verkerk T., Vermeulen W., Burgering B.M., de Zeeuw C.I., Grosveld F.,
RA Galjar N.;
RT "CLASPs are CLIP-115 and -170 associating proteins involved in the
RT regional regulation of microtubule dynamics in motile fibroblasts.";
RL Cell 104:923-935(2001).
DR EMBL; AJ237670; CAB92974.1; -.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_GLY; 2.
DR SMART; SM00343; ZnF_C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS50245; CAP_GLY_2; 2.
SQ SEQUENCE 1320 AA; 148264 MW; 87BE0F0463D20E9D CRC64;

Query Match 7.5%; Score 179.5; DB 11; Length 1320;
Best Local Similarity 22.6%; Pred. No. 0.62;
Matches 134; Conservative 77; Mismatches 211; Indels 171; Gaps 27;

```
Qy         22 KPGGSQDKSLH-NRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKA 80

Db        525 KPPGDVDMSLSLLQEISA---LQEKLEVTHTDHQNEVTSLKDHFGTREEMFQKEIKALHA 581

Qy         81 ITEKEKIEKERQSIRSSPLD--NKLNVEDV---------------------------- 108
```

```
                 |  ||: || :|:||  ||  || | : :
Db        582 AT--EKLSKENESLRSK-LDHANKENSDVIALWKSKLETAIASHQQAMEELKVSFSKGIG 638

Qy        109 -DS-------TKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTA---------EDI 151
                ||       |:  :|  ||       |  | ||     |  :  : |         ||
Db        639 TDSAEFAELKTQIERLRLDYQHEIESLQSK-QDSERSAHAKEMESMKAKLMKIIKEKEDS 697

Qy        152 VHKIAARIYEENDRAV--FDKIVSKLLNLGLITESQAHTLEDEVA------EVLQKLISK 203
                :  : ||:    |: :   ::::|||    :  |  |   |: |:      : : ||  |
Db        698 LEAVKARLDTAEDQHLVEMEEMLSKLQEAEIKKEKFASASEEAVSTQTSMQDTVNKLHQK 757

Qy        204 EANNYEEDPNKPTSWTENQAGKIPEKVTPMAA-------IQDGLAKG----END--ETV- 249
                  ||  |  :|  |    |: | :| | |        :| | |     |||  | :
Db        758 -----EEQFNMLSSELE----KLRENLTDMEAKFKEKDEREDQLVKAKEKLENDIAEIMK 808

Qy        250 -----SNTLTLTNGLERRTKTYSEDNFEELQY-----FPNFYALLKSI--------DSEK 291
                  |: ||  |  | | |   | : ||||         |   | |||         |::
Db        809 MSGDNSSQLTKMND-ELRLK---ERSVEELQLKLTKANENASLLQKSIGEVTLKAEQSQQ 864

Qy        292 EA--KEKETLITIMKTLIDFVKMM-----------VKYGTISPEEGVSY---LENLDEMI 335
                ||  | :|    :   |::  | |            ||   | |  : :   |:|  :|:
Db        865 EAAKKHEEEKKELENKLLELEKKMETSHYQCQDLKAKYEKASSETKIKHEEILQNFQKML 924

Qy        336 -------------------------------------ALQTKNKLEKNATDNISKL-- 354
                                                    |:|   :: |  |: :: |
Db        925 VDTEDKLKAAQEANRDLMQDMEELKSQADKAKAAQTAEDAMQIMEQMTKEKTETLASLED 984

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRK 414
              : |   | |: ||   | :|   |:   ::|     | :| | : |   :|  :
Db        985 TKQTNAKLQSELDTLKENNLKTVEELNKSKELLNEENQ----KMEEFKKEIETLKQAAAQ 1040

Qy        415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSS 467
              : |    ::  |   :| : || :        |: :|: :  |  | : |:
Db       1041 KSQQLSALQEENVKLAEELGRTRDEVTSHQKLEEERSVLNNQLLEMKKSLPSN 1093
```

Search completed: January 2, 2004, 12:35:46
Job time : 42 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 46 Seconds
(without alignments)
1614.870 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2399	100.0	468	21	AAB18926	A novel polypeptid
2	2399	100.0	468	22	AAU12216	Human PRO5990 poly
3	2399	100.0	468	22	AAB87600	Human PRO5990. Ho
4	2399	100.0	468	23	ABG95925	Human secreted/tra
5	2399	100.0	468	24	ABU69097	Human PRO polypept
6	2399	100.0	468	24	ABU69120	Human PRO polypept
7	2399	100.0	468	24	ABU71580	Human secreted pol
8	2399	100.0	468	24	ABU72026	Novel human secret
9	2399	100.0	468	24	ABU72183	Human PRO polypept
10	2399	100.0	468	24	ABU66614	Human PRO polypept
11	2399	100.0	468	24	ABU66890	Human secreted/tra
12	2399	100.0	468	24	ABU59695	Novel secreted and
13	2399	100.0	468	24	ABG73287	Human neurosecreto
14	2399	100.0	567	22	AAO13905	Human polypeptide
15	2392	99.7	468	22	AAM93910	Human polypeptide,
16	2384	99.4	468	21	AAY81955	Human HsgIII prote
17	2271	94.7	470	22	ABG12976	Novel human diagno
18	2102.5	87.6	471	23	ABB57046	Mouse ischaemic co
19	1974	82.3	396	24	AAO16322	Human secreted pro
20	594	24.8	316	22	ABG14006	Novel human diagno
21	475	19.8	212	22	ABG14003	Novel human diagno
22	184	7.7	665	21	AAB18278	Plasmodium falcipa
23	180.5	7.5	1881	23	ABP73809	Candida albicans e
24	179	7.5	1558	21	AAB18324	Plasmodium falcipa
25	179	7.5	1786	18	AAW24790	P. falciparum live
26	179	7.5	1787	23	AAU96699	Plasmodium falcipo
27	178.5	7.4	1392	20	AAY06999	Restin protein seq
28	175.5	7.3	1427	12	AAR10534	Human 160kD mediat
29	174.5	7.3	1164	19	AAW40537	Group B streptococ
30	173	7.2	1099	19	AAW40538	Mutant C-beta prot
31	172.5	7.2	1164	17	AAR85781	Group B Streptococ
32	172.5	7.2	1164	21	AAY84459	Amino acid sequenc
33	172.5	7.2	1979	21	AAB18171	Plasmodium falcipa
34	171	7.1	1093	19	AAW40540	Mutant C-beta prot
35	170	7.1	1213	22	AAM40016	Human polypeptide
36	169.5	7.1	1788	22	ABB62995	Drosophila melanog
37	169	7.0	1129	21	AAY84462	Amino acid sequenc
38	168	7.0	1164	19	AAW40541	Mutant C-beta prot
39	168	7.0	1164	21	AAY84463	Amino acid sequenc
40	167	7.0	493	13	AAR26944	P.falciparum LSA g
41	166.5	6.9	1135	21	AAY84460	Amino acid sequenc
42	166	6.9	1128	19	AAW40539	Mutant C-beta prot
43	166	6.9	1128	21	AAY84461	Amino acid sequenc
44	166	6.9	1192	21	AAB18165	Plasmodium falcipa
45	166	6.9	2633	22	ABG06505	Novel human diagno

ALIGNMENTS

RESULT 1
AAB18926
ID AAB18926 standard; Protein; 468 AA.
XX
AC AAB18926;
XX
DT 08-FEB-2001 (first entry)
XX
DE A novel polypeptide designated PRO5990.
XX
KW Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW insulinemia; kidney disorder; Bergers disease; nephropathy;
KW Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW Crohns disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 68..71
FT /note= "N-glycosylation site"
FT Modified-site 70..73
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 82..85
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 97..100
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 125..128
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 143..148

```
FT                   /note= "N-myristoylation site"
FT   Modified-site   147..150
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   188..191
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   217..220
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   239..244
FT                   /note= "N-myristoylation site"
FT   Modified-site   265..268
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   289..292
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   305..308
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   320..323
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   326..329
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   346..349
FT                   /note= "N-glycosylation site"
FT   Modified-site   350..353
FT                   /note= "N-glycosylation site"
FT   Modified-site   362..365
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   369..372
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   382..385
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   386..389
FT                   /note= "casein kinase II phosphorylation site"
FT   Modified-site   387..390
FT                   /note= "casein kinase II phosphorylation site"
XX
PN   WO200056889-A2.
XX
PD   28-SEP-2000.
XX
PF   01-MAR-2000; 2000WO-US05601.
XX
PR   23-MAR-1999;   99US-0125774.
PR   23-MAR-1999;   99US-0125778.
PR   24-MAR-1999;   99US-0125826.
PR   31-MAR-1999;   99US-0127035.
PR   05-APR-1999;   99US-0127706.
PR   21-APR-1999;   99US-0130359.
PR   27-APR-1999;   99US-0131270.
PR   27-APR-1999;   99US-0131272.
PR   27-APR-1999;   99US-0131291.
PR   04-MAY-1999;   99US-0132371.
PR   04-MAY-1999;   99US-0132379.
PR   04-MAY-1999;   99US-0132383.
PR   25-MAY-1999;   99US-0135750.
PR   08-JUN-1999;   99US-0138166.
PR   20-JUL-1999;   99US-0144791.
PR   03-AUG-1999;   99US-0146970.
PR   09-DEC-1999;   99US-0170262.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2000-628263/60.
DR   N-PSDB; AAA96353.
XX
PT   Novel secreted and transmembrane polypeptides useful for diagnosing
PT   tumour in a mammal, for identifying agonists and antagonists of the
PT   polypeptide and for therapeutic use  -
XX
PS   Claim 12; Fig 36; 222pp; English.
XX
CC   The present sequence represents a secreted or transmembrane polypeptide.
CC   The specification describes polypeptides designated PRO1484, PRO4334,
CC   PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC   PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC   PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC   useful for diagnosing tumour in a mammal. The polypeptides, their
CC   agonists and antagonists are useful treating a condition associated with
CC   expression or activity of the polypeptide. Conditions treated include
CC   obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are
CC   capable of inducing proliferation of mammalian kidney mesangial cells
CC   and are therefore useful for treating kidney disorders associated with
CC   decreased mesangial cell function such as Bergers disease or other
CC   nephropathies associated with Schonlein-Henoch purpura, celiac disease,
CC   dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
CC   to generate transgenic animals for use in development and screening of
CC   therapeutically useful reagents and also for chromosome identification
CC   and tissue typing.
XX
SQ   Sequence   468 AA;

  Query Match            100.0%;  Score 2399;  DB 21;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 2
AAU12216
ID   AAU12216 standard; Protein; 468 AA.
XX
AC   AAU12216;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO5990 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI   Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI   Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21288.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical -
XX
PS   Claim 12; Fig 90; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   468 AA;
```

```
  Query Match           100.0%;  Score 2399;  DB 22;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

```
RESULT 3
AAB87600
ID   AAB87600 standard; Protein; 468 AA.
XX
AC   AAB87600;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO5990.
XX
KW   Human; PRO protein; mapping.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
```

```
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92132.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 150; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the PRO protein,
CC   agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC   employed as molecular weight markers for protein electrophoresis. The PRO
CC   coding sequence has applications in molecular biology, including use as
CC   hybridisation probes, and in chromosome and gene mapping.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 22;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 4
ABG95925
ID   ABG95925 standard; Protein; 468 AA.
XX
AC   ABG95925;
XX
DT   10-DEC-2002  (first entry)
XX
DE   Human secreted/transmembrane protein PRO5990.
XX
KW   Human; secreted protein; transmembrane protein; antirheumatic;
KW   antiarthritic; osteopathic; sports-related joint problem;
KW   articular cartilage defect; osteoarthritis; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
PN   US2002119130-A1.
XX
PD   29-AUG-2002.
XX
PF   06-DEC-2001; 2001US-0006867.
XX
PR   29-OCT-1997;   97US-063435P.
PR   29-OCT-1997;   97US-064215P.
PR   22-APR-1998;   98US-082797P.
PR   29-APR-1998;   98US-083495P.
PR   15-MAY-1998;   98US-085579P.
PR   10-JUN-1998;   98US-088811P.
PR   10-JUN-1998;   98US-088824P.
PR   10-JUN-1998;   98US-088825P.
PR   11-JUN-1998;   98US-088863P.
PR   12-JUN-1998;   98US-089105P.
PR   16-JUN-1998;   98US-089514P.
PR   16-SEP-1998;   98WO-US19330.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21194.
PR   22-DEC-1999;   99WO-US30720.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   30-MAR-2000; 2000WO-US08439.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   10-NOV-2000; 2000WO-US30873.
PR   01-DEC-2000; 2000WO-US32378.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2002-731348/79.
DR   N-PSDB; ABS74452.
XX
PT   New isolated secreted and transmembrane PRO polypeptide useful for
PT   modulating biological activity of a cell, or for treating
PT   sports-related joint problems, osteoarthritis or rheumatoid arthritis
XX
PS   Claim 20; Fig 150; 399pp; English.
XX
CC   The invention relates to an isolated secreted and transmembrane PRO
CC   polypeptide having 80 % sequence identity to a sequence appearing
CC   as ABG95851-ABG95934 or their associated signal peptide, or a sequence of
CC   an extracellular domain of the proteins with their associated signal
CC   peptide or lacking its associated signal peptide. Also included are
CC   the nucleic acids encoding the proteins, vectors, host cells,
CC   fusion proteins and antibodies which specifically bind to the proteins.
CC   The proteins are useful for detecting a polypeptide designated as A, B, C
CC   or D in a sample suspected of containing an A, B, C or D polypeptide,
CC   by contacting the sample with a polypeptide designated as E, F, G, H or
CC   I (or vice versa) and determining the formation of a A/E, B/F, B/G, C/H
CC   or D/I polypeptide conjugate in the sample, where the formation of the
CC   conjugate is indicative of the presence of an A, B, C or D polypeptide
CC   in the sample, where A is a PRO10272 polypeptide, B is a PRO20110
CC   polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,
CC   E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040
```

```
CC   polypeptide, H is a PRO20233 polypeptide and I is a PRO1890
CC   polypeptide. The sample comprises a cell suspected of expressing the A,
CC   B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with
CC   a detectable label or is attached to a solid support. The proteins are
CC   useful for linking a bioactive molecule to a cell expressing a
CC   polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive
CC   molecule is a toxin, a radiolabel or an antibody. The bioactive molecule
CC   causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies
CC   against them are useful for modulating a biological activity of a cell
CC   expressing a polypeptide designated as A, B, C or D or E, F, G, H, or
CC   I. The cell is killed. The proteins are useful for identifying
CC   agonists or antagonists, for the preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the proteins, as
CC   molecular weight markers for protein electrophoresis purposes, and as
CC   therapeutic agents for treating sports-related joint problems,
CC   articular cartilage defects, osteoarthritis or rheumatoid arthritis.
CC   Nucleic acids encoding the proteins are useful as hybridisation probes,
CC   in chromosome and gene mapping, in the generation of anti-sense RNA and
CC   DNA, for the preparation of the proteins, to generate transgenic or
CC   knockout animals which are useful in the development and screening of
CC   therapeutic useful reagents, for chromosome identification, and in gene
CC   therapy. The antibody is useful as a therapeutic agent, in a diagnostic
CC   assay and for affinity purification of the protein from recombinant
CC   cell culture natural sources. The present sequence represents a novel
CC   secreted or transmembrane protein of the invention.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 23;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 5
ABU69097
ID   ABU69097 standard; Protein; 468 AA.
XX
AC   ABU69097;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human PRO polypeptide #18.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein; cancer;
KW   non-insulin dependent diabetes mellitus; septic shock; stroke;
KW   rheumatoid arthritis; graft-versus-host disease; cardiac ischaemia;
KW   psoriasis; inflammatory bowel disease; asthma; antidiabetic;
KW   cytostatic; immunosuppressive; antirheumatic; antiarthritic;
KW   cerebroprotective; vasotropic; antipsoriatic; antiinflammatory;
KW   antiasthmatic.
XX
OS   Homo sapiens.
XX
PN   US2003008348-A1.
XX
PD   09-JAN-2003.
XX
PF   26-DEC-2001; 2001US-0035855.
XX
PR   14-MAY-1999;   99WO-US10733.
PR   02-DEC-1999;   99WO-US28551.
PR   22-DEC-1999;   99WO-US30720.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   15-MAY-1998;   98US-085579P.
PR   15-DEC-1998;   98US-112514P.
PR   22-DEC-1998;   98US-113300P.
PR   23-DEC-1998;   98US-113430P.
PR   23-DEC-1998;   98US-113605P.
PR   22-JAN-1999;   99US-116843P.
PR   23-MAR-1999;   99US-125774P.
PR   23-MAR-1999;   99US-125778P.
PR   24-MAR-1999;   99US-125826P.
PR   31-MAR-1999;   99US-127035P.
PR   05-APR-1999;   99US-127706P.
PR   13-APR-1999;   99US-129122P.
PR   21-APR-1999;   99US-130359P.
PR   25-MAY-1999;   99US-135750P.
PR   08-JUN-1999;   99US-138166P.
PR   20-JUL-1999;   99US-144791P.
PR   03-AUG-1999;   99US-146970P.
PR   29-OCT-1999;   99US-162506P.
PR   16-AUG-2001; 2001US-0931836.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2003-341326/32.
DR   N-PSDB; ACA06116.
XX
PT   New PRO polypeptides and nucleic acid molecules, useful for diagnosing
PT   or treating diabetes mellitus, cancers, septic shock, inflammatory
PT   bowel disease or asthma, or in gene therapy, chromosome identification
PT   or tissue typing -
XX
PS   Claim 12; Fig 36; 196pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful in diagnosing or
CC   treating non-insulin dependent diabetes mellitus, cancers, septic
CC   shock, rheumatoid arthritis, graft-versus-host disease, stroke, cardiac
```

```
CC   ischaemia, psoriasis, inflammatory bowel disease or asthma. The PRO
CC   polynucleotide sequences may be used as hybridisation probes in
CC   chromosome and gene mapping, or in generating antisense RNA and DNA.
CC   They are also useful in preparing PRO polypeptides, in assays to
CC   identify other proteins or molecules involved in binding reaction, to
CC   generate transgenic animals or knockout animals, which in turn are
CC   useful in the development and screening of therapeutically useful
CC   reagents, for chromosome identification, and tissue typing. The PRO
CC   polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC   PRO polypeptides, or for the affinity purification of the polypeptides
CC   from recombinant cell culture or natural sources. ABU69080-ABU69102
CC   represent the human PRO polypeptides of the invention.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 6
ABU69120
ID   ABU69120 standard; Protein; 468 AA.
XX
AC   ABU69120;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human PRO polypeptide #18.
XX
KW   Human; secreted and transmembrane protein; bone disorder; obesity;
KW   cartilage disorder; sports injury; arthritis; diabetes mellitus;
KW   hypo-insulinaemia; obesity; hyper-insulinaemia; thalassaemia;
KW   haemoglobin-associated disorder; kidney disorder; Berger disease;
KW   mesangial cell function; nephropathy; Schonlein-Henoch purpura;
KW   celiac disease; dermatitis herpetiformis; Crohn's disease;
KW   anorectic; antiarthritic; antidiabetic; antianaemic; nephrotropic;
KW   antiinflammatory.
XX
OS   Homo sapiens.
XX
PN   US2003032061-A1.
XX
PD   13-FEB-2003.
XX
PF   26-DEC-2001; 2001US-0036214.
XX
PR   14-MAY-1999;   99WO-US10733.
PR   02-DEC-1999;   99WO-US28551.
PR   22-DEC-1999;   99WO-US30720.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   15-MAY-1998;   98US-085579P.
PR   15-DEC-1998;   98US-112514P.
PR   22-DEC-1998;   98US-113300P.
PR   23-DEC-1998;   98US-113430P.
PR   23-DEC-1998;   98US-113605P.
PR   23-DEC-1998;   98US-113621P.
PR   23-DEC-1998;   98US-114140P.
PR   12-JAN-1999;   99US-115552P.
PR   22-JAN-1999;   99US-116843P.
PR   23-MAR-1999;   99US-125774P.
PR   23-MAR-1999;   99US-125778P.
PR   24-MAR-1999;   99US-125826P.
PR   31-MAR-1999;   99US-127035P.
PR   05-APR-1999;   99US-127706P.
PR   13-APR-1999;   99US-129122P.
PR   21-APR-1999;   99US-130359P.
PR   27-APR-1999;   99US-131270P.
PR   27-APR-1999;   99US-131272P.
PR   27-APR-1999;   99US-131291P.
PR   04-MAY-1999;   99US-132371P.
PR   04-MAY-1999;   99US-132379P.
PR   04-MAY-1999;   99US-132383P.
PR   25-MAY-1999;   99US-135750P.
PR   08-JUN-1999;   99US-138166P.
PR   20-JUL-1999;   99US-144791P.
PR   03-AUG-1999;   99US-146970P.
PR   29-OCT-1999;   99US-162506P.
PR   16-AUG-2001; 2001US-0931836.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2003-341962/32.
DR   N-PSDB; ACA06173.
XX
PT   Novel isolated PRO polypeptides e.g., PRO4334, PRO1122, PRO1889,
PT   PRO1890, PRO1887, PRO1785, PRO4353, useful for treating sports
PT   injuries, arthritis, diabetes, obesity, hyper- or hypo-insulinaemia -
XX
PS   Claim 12; Fig 36; 194pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful in diagnosing or
CC   treating various bone and/or cartilage disorders (e.g. sports
CC   injuries, arthritis), various insulin deficient states (e.g. diabetes
```

```
CC   mellitus, hypo-insulinaemia), obesity, hyper-insulinaemia,
CC   haemoglobin-associated disorders (e.g. thalassaemias), kidney disorders
CC   associated with decreased mesangial cell function (e.g. Berger disease),
CC   or other nephropathies associated with Schonlein-Henoch purpura,
CC   celiac disease, dermatitis herpetiformis or Crohn's disease. The PRO
CC   polynucleotide sequences may be used as hybridisation probes in
CC   chromosome and gene mapping, or in generating antisense RNA and DNA.
CC   They are also useful in preparing PRO polypeptides, in assays to
CC   identify other proteins or molecules involved in binding reaction, to
CC   generate transgenic animals or knockout animals, which in turn are
CC   useful in the development and screening of therapeutically useful
CC   reagents, for chromosome identification, and tissue typing. The PRO
CC   polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. Anti-PRO antibodies may be used in diagnostic assays for
CC   PRO polypeptides, or for the affinity purification of the polypeptides
CC   from recombinant cell culture or natural sources. ABU69103-ABU69125
CC   represent the human PRO polypeptides of the invention.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 7
ABU71580
ID   ABU71580 standard; Protein; 468 AA.
XX
AC   ABU71580;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human secreted polypeptide PRO5990.
XX
KW   Human; gene therapy; tumour; cancer.
XX
OS   Homo sapiens.
XX
PN   US2003013855-A1.
XX
PD   16-JAN-2003.
XX
PF   03-MAY-2002; 2002US-0063616.
XX
PR   30-DEC-1998;   98KR-0062142.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   30-DEC-1999;   99WO-US31274.
PR   18-FEB-2000; 2000WO-US04341.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   21-MAR-2000; 2000WO-US07532.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   24-AUG-2000; 2000WO-US23328.
PR   10-NOV-2000; 2000WO-US30873.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   14-MAY-1999;   99US-0311832.
PR   25-AUG-1999;   99US-0380137.
PR   25-AUG-1999;   99US-0380138.
PR   25-AUG-1999;   99US-0380139.
PR   25-AUG-1999;   99US-0380142.
PR   15-SEP-1999;   99US-0397342.
PR   18-OCT-1999;   99US-0403297.
PR   12-NOV-1999;   99US-0423844.
PR   22-AUG-2000; 2000US-0644848.
PR   18-SEP-2000; 2000US-0664610.
PR   18-SEP-2000; 2000US-0665350.
PR   08-NOV-2000; 2000US-0709238.
PR   20-DEC-2000; 2000US-0747259.
PR   22-MAR-2001; 2001US-0816744.
PR   10-MAY-2001; 2001US-0854208.
PR   10-MAY-2001; 2001US-0854280.
PR   30-MAY-2001; 2001US-0870574.
PR   05-JUN-2001; 2001US-0874503.
PR   29-JUN-2001; 2001US-0869599.
PR   18-JUL-2001; 2001US-0908827.
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-330485/31.
DR   N-PSDB; ACA58884.
XX
PT   New isolated antibody specifically binding a PRO polypeptide, useful
PT   for the preparation of a medicament for treating disorders with the
PT   aberrant expression or activity of the PRO polypeptide, such as tumor
PT   conditions and cancer  -
XX
PS   Example 17; Page 218-219; 406pp; English.
XX
CC   The invention relates to an antibody that binds to a polypeptide with a
CC   fully defined sequence given in the specification. The methods and
CC   compositions (containing antibodies that specifically bind a PRO
CC   polypeptide) of the present invention are useful for the preparation of a
CC   medicament for the treatment of disorders associated with the aberrant
CC   expression or activity of the PRO polypeptide, such as tumour conditions
CC   and cancer. They can also be used to generate transgenic or knockout
CC   animals useful in the development and screening of therapeutically useful
CC   reagents. The PRO polypeptides and encoding nucleic acids can be used as
CC   molecular weight markers for protein electrophoresis, chromosome
CC   identification and tissue typing. The PRO polypeptides are useful to
CC   induce angiogenesis e.g wound healing; in the treatment of sports-related
CC   joint problems, articular cartilage defects, osteoarthritis or rheumatoid
CC   arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The
CC   antibodies may be used in various diagnostic, competitive binding and/or
```

```
CC   immunoprecipitation assays. The present sequence represents the amino
CC   acid sequence of a PRO polypeptide of the invention.
XX
SQ   Sequence   468 AA;

  Query Match          100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 8
ABU72026
ID   ABU72026 standard; Protein; 468 AA.
XX
AC   ABU72026;
XX
DT   11-JUN-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO5990.
XX
KW   Human; secreted and transmembrane polypeptide;
KW   chromosome mapping; gene mapping; transgenic animal; knockout animal;
KW   therapeutic agent screening; chromosome identification; tissue typing;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   US2003018183-A1.
XX
PD   23-JAN-2003.
XX
PF   01-MAY-2002; 2002US-0063512.
XX
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-330984/31.
DR   N-PSDB; ACA60437.
XX
PT   New secreted and transmembrane PRO polypeptides and nucleic acid
PT   molecules encoding the polypeptides, useful in gene therapy or
PT   preparing a medicament for treating a condition that is responsive to
PT   the PRO polypeptide or antibody  -
XX
PS   Disclosure; Fig 150; 409pp; English.
XX
CC   The invention describes novel isolated PRO polypeptides. The PRO
CC   polypeptides or anti-PRO antibodies are useful in preparing a medicament
CC   for treating a condition that is responsive to the PRO polypeptide or
CC   antibody. The PRO nucleotide sequences may be used as hybridisation
CC   probes in chromosome and gene mapping, or in generating antisense RNA
CC   and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,
CC   in assays to identify other proteins or molecules involved in binding
CC   reaction, to generate transgenic animals or knockout animals, which in
CC   turn are useful in the development and screening of therapeutically
CC   useful reagents, for chromosome identification, and tissue typing. The
CC   PRO polypeptides and nucleic acid molecules are also useful in gene
CC   therapy, and as molecular weight markers for protein electrophoresis
CC   purposes. The anti-PRO antibodies may be used in diagnostic assays for
CC   PRO, or for the affinity purification of PRO from recombinant cell
CC   culture or natural sources. This is the amino acid sequence of a novel
CC   human secreted and transmembrane PRO poypeptide.
XX
SQ   Sequence   468 AA;

  Query Match          100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 9
ABU72183
ID   ABU72183 standard; Protein; 468 AA.
XX
AC   ABU72183;
XX
DT   13-JUN-2003  (first entry)
XX
```

```
DE   Human PRO polypeptide #75.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein;
KW   anti-PRO antibody; diagnostic assay; gene expression.
XX
OS   Homo sapiens.
XX
PN   US2003023042-A1.
XX
PD   30-JAN-2003.
XX
PF   01-MAY-2002; 2002US-0063502.
XX
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-331484/31.
DR   N-PSDB; ACA63447.
XX
PT   Novel monoclonal antibody that binds to secreted and transmembrane
PT   polypeptide, useful for detecting and purifying the polypeptide and
PT   also for treating conditions responsive to the antibody  -
XX
PS   Disclosure; Fig 150; 408pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful for preparing a
CC   medicament useful in the treatment of a condition responsive to
CC   anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic
CC   assays for PRO, by detecting its expression in specific cells,
CC   tissues or serum, and for affinity purification of PRO from
CC   recombinant cell culture or natural sources. ABU72109-ABU72192
CC   represent the human PRO polypeptides of the invention.
XX
SQ   Sequence   468 AA;

  Query Match            100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 10
ABU66614
ID   ABU66614 standard; Protein; 468 AA.
XX
AC   ABU66614;
XX
DT   23-MAY-2003  (first entry)
XX
DE   Human PRO polypeptide #45.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein;
KW   tumour necrosis factor-alpha; TNF-alpha; blood; proliferation;
KW   differentiation; chondrocyte; tumour; genetic disorder;
KW   cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003036180-A1.
XX
PD   20-FEB-2003.
XX
PF   09-MAY-2002; 2002US-0143114.
XX
PR   31-MAR-1997;   97WO-US05230.
PR   12-JUN-1998;   98WO-US12456.
PR   14-JUL-1998;   98WO-US14552.
PR   28-AUG-1998;   98WO-US17888.
PR   10-SEP-1998;   98WO-US18824.
PR   14-SEP-1998;   98WO-US19093.
PR   14-SEP-1998;   98WO-US19094.
PR   14-SEP-1998;   98WO-US19177.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   29-OCT-1998;   98WO-US22991.
PR   29-OCT-1998;   98WO-US22992.
PR   20-NOV-1998;   98WO-US24855.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   10-MAR-1999;   99WO-US05190.
PR   20-APR-1999;   99WO-US08615.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   22-DEC-1999;   99WO-US30720.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
```

```
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05746.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   11-AUG-2000; 2000WO-US22031.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   10-NOV-2000; 2000WO-US30873.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-MAR-2001; 2001WO-US06666.
PR   25-MAY-2001; 2001WO-US17092.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   22-JUN-2001; 2001WO-US20116.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   20-DEC-2000; 2000US-0747259.
PR   28-FEB-2001; 2001US-0796498.
PR   09-MAR-2001; 2001US-0802706.
PR   14-MAR-2001; 2001US-0808689.
PR   22-MAR-2001; 2001US-0816744.
PR   05-APR-2001; 2001US-0828366.
PR   10-MAY-2001; 2001US-0854208.
PR   10-MAY-2001; 2001US-0854280.
PR   18-MAY-2001; 2001US-0860216.
PR   25-MAY-2001; 2001US-0866028.
PR   25-MAY-2001; 2001US-0866034.
PR   01-JUN-2001; 2001US-0872035.
PR   05-JUN-2001; 2001US-0874503.
PR   14-JUN-2001; 2001US-0882636.
PR   19-JUN-2001; 2001US-0886342.
PR   21-JUN-2001; 2001US-0887879.
PR   18-JUL-2001; 2001US-0908827.
PR   06-AUG-2001; 2001US-0924419.
PR   09-AUG-2001; 2001US-0927796.
PR   16-AUG-2001; 2001US-0931836.
PR   19-DEC-2001; 2001US-0028072.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
PI   Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI   Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2003-332040/31.
DR   N-PSDB; ACA03647.
XX
PT   New secreted and transmembrane PRO nucleic acids, useful for gene
PT   therapy, in chromosome and gene mapping, as chromosome markers, in
PT   tissue typing, and in chromosome identification -
XX
PS   Claim 12; Fig 90; 660pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides are useful for detecting other PRO polypeptides, for
CC   linking bioactive molecules to cells expressing PRO polypeptides,
CC   for modulating biological activities of cells expressing PRO
CC   polypeptides, and for for identifying agonists or antagonists.
CC   The PRO polypeptides are useful for for stimulating the release of
CC   tumour necrosis factor (TNF)-alpha from human blood, for stimulating
CC   the proliferation or differentiation of chondrocytes, and detecting the
CC   presence of tumours. The polynucleotide sequences encoding PRO
CC   polypeptides are useful as hybridisation probes, in chromosome and
CC   gene mapping, in the generation of antisense RNA and DNA, in the
CC   preparation of PRO polypeptides, for generating transgenic animals or
CC   knockout animals, for the genetic analysis of individuals with genetic
CC   disorders, and in gene therapy. ABU66570-ABU66844 represent the human
CC   PRO polypeptides of the invention.
CC   Note: The sequence data for this patent was obtained in electronic
CC   format directly from the USPTO web site at
CC   seqdata.uspto.gov/psipsDIDEntry.html.
XX
SQ   Sequence   468 AA;

  Query Match            100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 11
ABU66890
ID   ABU66890 standard; Protein; 468 AA.
XX
AC   ABU66890;
XX
DT   27-MAY-2003  (first entry)
XX
DE   Human secreted/transmembrane, PRO, protein SEQ ID 90.
XX
KW   Human; secreted protein; transmembrane protein; PRO;
KW   inflammatory disease; organ failure; atherosclerosis; cardiac injury;
KW   infertility; birth defects; premature aging; AIDS; biosensor;
KW   acquired immunodeficiency syndrome; cancer; diabetic complication;
KW   bioreactor; tumour.
```

XX
OS Homo sapiens.
XX
PN US2003032155-A1.
XX
PD 13-FEB-2003.
XX
PF 03-MAY-2002; 2002US-0137865.
XX
PR 31-MAR-1997; 97WO-US05230.
PR 12-JUN-1998; 98WO-US12456.
PR 14-JUL-1998; 98WO-US14552.
PR 28-AUG-1998; 98WO-US17888.
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19093.
PR 14-SEP-1998; 98WO-US19094.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 29-OCT-1998; 98WO-US22991.
PR 29-OCT-1998; 98WO-US22992.
PR 20-NOV-1998; 98WO-US24855.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 99WO-US00106.
PR 08-MAR-1999; 99WO-US05028.
PR 10-MAR-1999; 99WO-US05190.
PR 20-APR-1999; 99WO-US08615.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 22-DEC-1999; 99WO-US30720.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05746.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 25-MAY-2001; 2001WO-US17092.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 22-JUN-2001; 2001WO-US20116.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 20-DEC-2000; 2000US-0747259.
PR 28-FEB-2001; 2001US-0796498.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 18-MAY-2001; 2001US-0860216.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 01-JUN-2001; 2001US-0872035.
PR 05-JUN-2001; 2001US-0874503.
PR 14-JUN-2001; 2001US-0882636.
PR 19-JUN-2001; 2001US-0886342.
PR 21-JUN-2001; 2001US-0887879.
PR 18-JUL-2001; 2001US-0908827.
PR 06-AUG-2001; 2001US-0924419.
PR 09-AUG-2001; 2001US-0927796.
PR 16-AUG-2001; 2001US-0931836.
PR 19-DEC-2001; 2001US-0028072.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2003-331925/31.
DR N-PSDB; ACA04068.
XX
PT New secreted and transmembrane nucleic acids and polypeptides,
PT designated as PRO, useful for treating inflammation, organ failure,
PT atherosclerosis, cardiac injury, infertility, birth defects, premature
PT aging, AIDS, or cancer -
XX
PS Claim 12; Fig 90; 659pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising, or which is
CC at least 80% identical to, or the full-length coding sequence of, any of
CC the 275 nucleotide sequences, encoding the corresponding PRO polypeptide
CC (one of 275 secreted or transmembrane proteins). The nucleic acid
CC further comprises the full-length coding sequence of the DNA deposited
CC under American Type Culture Collection (ATCC) accession number in a list
CC given in the specification. Also included are vectors and host
CC cells for producing PRO proteins, PRO fusion proteins, anti-PRO
CC antibodies, PRO extracellular domains and mature sequences, methods
CC of detecting PRO proteins, methods for stimulating the release of
CC TNF-alpha (tumour necrosis factor alpha) from human blood,
CC (and the proliferation of differentiation of chondrocyte cells, the
CC proliferation of, or gene expression in pericyte cells, the release or
CC proteoglycans from cartilage, proliferation of inner ear urticular
CC supporting cells, the proliferation of T-lymphocyte cells, the release
CC of a cytokine from peripheral blood mononuclear cells (PBMC), or the
CC proliferation of endothelial cells), a method for modulating the uptake
CC of glucose or free fatty acid (FFA) by skeletal muscle cells,
CC a method for inhibiting the binding of A-peptide to factor VIIA,
CC or the differentiation of adipocyte cells, a method for detecting the
CC presence of a tumour in a mammal and an oligonucleotide probe derived

```
CC   from any of the nucleotide sequences cited above. The nucleic acids and
CC   polypeptides are useful for treating inflammatory diseases, organ
CC   failure, atherosclerosis, cardiac injury, infertility, birth defects,
CC   premature aging, AIDS (acquired immunodeficiency syndrome), cancer, or
CC   diabetic complications. The nucleic acids are useful as hybridisation
CC   probes, in chromosome and gene mapping, and in generating antisense RNA
CC   or DNA. The polypeptides are useful as pharmaceuticals, diagnostics,
CC   biosensors or bioreactors. Both are useful in tissue typing.
CC   The present sequence represents a PRO protein of the invention.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 12
ABU59695
ID   ABU59695 standard; Protein; 468 AA.
XX
AC   ABU59695;
XX
DT   13-MAY-2003  (first entry)
XX
DE   Novel secreted and transmembrane protein PRO5990.
XX
KW   Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;
KW   cardiac insufficiency disorder; cancer; tumour; immune response;
KW   adrenal cortical capillary endothelial growth; c-fos induction;
KW   vascular endothelial growth factor inhibition; VEGF inhibition;
KW   endothelial cell growth inhibitor; T-lymphocytes stimulation;
KW   retinal neurons cell survival; rod photoreceptor cell survival;
KW   retinal disorder; retinitis pigmentosum; kidney disorder;
KW   mammalian kidney mesangial cell proliferation; Berger disease;
KW   dermatitis; herpetiformis; Crohn's disease; chondrocyte proliferation;
KW   chondrocyte redifferentiation; sports injury; arthritis.
XX
OS   Homo sapiens.
XX
PN   US2003017563-A1.
XX
PD   23-JAN-2003.
XX
PF   07-MAY-2002; 2002US-0140808.
XX
PR   31-MAR-1997;   97WO-US05230.
PR   12-JUN-1998;   98WO-US12456.
PR   14-JUL-1998;   98WO-US14552.
PR   28-AUG-1998;   98WO-US17888.
PR   10-SEP-1998;   98WO-US18824.
PR   14-SEP-1998;   98WO-US19093.
PR   14-SEP-1998;   98WO-US19094.
PR   14-SEP-1998;   98WO-US19177.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   29-OCT-1998;   98WO-US22991.
PR   29-OCT-1998;   98WO-US22992.
PR   20-NOV-1998;   98WO-US24855.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   10-MAR-1999;   99WO-US05190.
PR   20-APR-1999;   99WO-US08615.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   22-DEC-1999;   99WO-US30720.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05746.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   11-AUG-2000; 2000WO-US22031.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   10-NOV-2000; 2000WO-US30873.
```

PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 25-MAY-2001; 2001WO-US17092.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 22-JUN-2001; 2001WO-US20116.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 20-DEC-2000; 2000US-0747259.
PR 28-FEB-2001; 2001US-0796498.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 18-MAY-2001; 2001US-0860216.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 01-JUN-2001; 2001US-0872035.
PR 05-JUN-2001; 2001US-0874503.
PR 14-JUN-2001; 2001US-0882636.
PR 19-JUN-2001; 2001US-0886342.
PR 21-JUN-2001; 2001US-0887879.
PR 18-JUL-2001; 2001US-0908827.
PR 06-AUG-2001; 2001US-0924419.
PR 09-AUG-2001; 2001US-0927796.
PR 16-AUG-2001; 2001US-0931836.
PR 19-DEC-2001; 2001US-0028072.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2003-148238/14.
DR N-PSDB; ABX89185.
XX
PT Novel isolated PRO polypeptides e.g., PRO826, PRO1068, PRO1184, PRO1346
PT and PRO1375, which stimulate proliferation of stimulated T-lymphocytes
PT are therapeutically useful for enhancing immune response and in cancer
PT treatments -
XX
PS Claim 12; Fig 90; 659pp; English.
XX
CC The invention describes an isolated human PRO polypeptide. The PRO
CC polypeptides are useful in detecting PRO polypeptides in a sample, in
CC linking a bioactive molecule to a cell expressing a PRO polypeptide, and
CC in modulating at least one biological activity of a cell expressing a PRO
CC polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus
CC useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186
CC stimulate adrenal cortical capillary endothelial growth, and PRO536,
CC PRO943, PRO828, PRO826, PRO1068 or PRO535, PRO826, PRO819, PRO1126,
CC PRO1360 and PRO1387 induce c-fos in endothelial cells, and are thus
CC useful for treating conditions or disorders where angiogenesis would be
CC beneficial, e.g. wound healing and antagonist of this polypeptide are
CC useful for treating cancerous tumours. PRO812 inhibits vascular
CC endothelial growth factor (VEGF) stimulated proliferation of endothelial
CC cells and is thus useful for inhibiting endothelial cell growth in
CC mammals which would be beneficial in inhibiting tumour growth. PRO826,
CC PRO1068, PRO1184, PRO1346 and PRO1375 stimulate proliferation of
CC stimulated T-lymphocytes and are therapeutically useful for enhancing
CC immune response. PRO828, PRO826, PRO1068 or PRO1132 enhance survival of
CC retinal neurons cells (PRO1132 is also enhances survival/proliferation of
CC rod photoreceptor cells) and therefore are useful for treating retinal
CC disorders of injuries, e.g. retinitis pigmentosum, AMD. PRO819, PRO813
CC and PRO11066 induce proliferation of mammalian kidney mesangial cells,
CC and therefore are useful for treating kidney disorders associated with
CC decreased mesangial cell function such as Berger disease or other
CC nephropathies associated with dermatitis, herpetiformis or Crohn's
CC disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the
CC proliferation and/or redifferentiation of chondrocytes in culture and
CC are thus useful for treating sports injuries, and arthritis. This
CC is the amino acid sequence of a novel human PRO protein.
XX
SQ Sequence 468 AA;

Query Match 100.0%; Score 2399; DB 24; Length 468;
Best Local Similarity 100.0%; Pred. No. 4.1e-146;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 13
ABG73287
ID ABG73287 standard; Protein; 468 AA.
XX
AC ABG73287;
XX
DT 28-APR-2003 (first entry)
XX
DE Human neurosecretory protein-1 (HUNSP-1).
XX
KW Human; neurosecretory protein-1; HUNSP-1; neurological disorder;
KW epilepsy; Huntington's disease; dementia; Alzheimer's disease;
KW multiple sclerosis; Parkinson's disease; schizophrenia; amnesia;
KW endocrine disorder; hypogonadism; Sheehan's syndrome; leukaemia;
KW Hashimoto's disease; Grave's disease; Cushing's disease; cancer;
KW adenocarcinoma; diabetes insipidus; melanoma; myeloma sarcoma;
KW bone cancer; breast cancer; cancer of the pancreas; brain cancer;
KW prostate cancer; skin cancer; thyroid cancer; antidiabetic; nootropic;
KW neuroprotective; neuroleptic; cytostatic; anticonvulsant; thyromimetic;
KW lymphoma; antiparkinsonian; antithyroid; immunosuppressive; antithyroid;
KW pancreatic islet cell cDNA library; ISLTNOT01; Incyte clone 2379427.
XX
OS Homo sapiens.
XX
PN US2002156014-A1.
XX
PD 24-OCT-2002.
XX
PF 14-FEB-2002; 2002US-0077040.

```
XX
PR   17-APR-1998;   98US-0062601.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P,  Corley NC,  Patterson C;
XX
DR   WPI; 2003-238117/23.
DR   N-PSDB; ABX11115.
XX
PT   New human neurosecretory proteins (HUNSP) and polynucleotides, useful
PT   for diagnosing, treating or preventing e.g. epilepsy, Alzheimer's
PT   disease, schizophrenia, Grave's disease, leukaemia, lymphoma, or breast
PT   or brain cancer -
XX
PS   Claim 1; Fig 1; 45pp; English.
XX
CC   The present invention relates to the isolation of human neurosecretory
CC   proteins -1 and -2 (HUNSP-1 and HUNSP-2), and the polynucleotide
CC   sequences encoding them. The HUNSP polypeptides, polynucleotides, and
CC   agonists are useful for treating a disease or condition associated with
CC   decreased expression of functional HUNSP. Antagonists of HUNSP are
CC   useful for treating a disease or condition associated with
CC   overexpression of functional HUNSP. HUNSP antibodies are useful for
CC   diagnosing a condition or disease associated with the expression of
CC   HUNSP. The sequences of the invention are particularly useful for
CC   diagnosing, treating or preventing neurological disorders
CC   (e.g. epilepsy, Huntington's disease, dementia, Alzheimer's disease,
CC   multiple sclerosis, Parkinson's disease, schizophrenia, amnesia),
CC   endocrine disorders (e.g. hypogonadism, Sheehan's syndrome, diabetes
CC   insipidus, Hashimoto's disease, Grave's disease, Cushing's disease),
CC   and cancers (e.g. adenocarcinoma, leukaemia, lymphoma, melanoma,
CC   myeloma sarcoma, bone cancer, breast cancer, cancer of the pancreas,
CC   brain cancer, prostate cancer, skin cancer or thyroid cancer).
CC   The present sequence represents HUNSP-1. The cDNA encoding HUNSP-1
CC   is isolated from pancreatic islet cell cDNA library (ISLTNOT01) Incyte
CC   clone 2379427.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 2399;  DB 24;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

```
RESULT 14
AAO13905
ID   AAO13905 standard; Protein; 567 AA.
XX
AC   AAO13905;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 27797.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI93836.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 27797; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   567 AA;

  Query Match             100.0%;  Score 2399;  DB 22;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-146;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 159

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 219

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 279

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 339

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 399

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 459

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 519

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
                 ||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 567
```

RESULT 15
AAM93910
ID AAM93910 standard; Protein; 468 AA.
XX
AC AAM93910;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 4061.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94872.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 4061; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX
SQ Sequence 468 AA;

Query Match 99.7%; Score 2392; DB 22; Length 468;
Best Local Similarity 99.8%; Pred. No. 1.2e-145;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEGIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

Search completed: January 2, 2004, 11:01:00
Job time : 48 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:32 ; Search time 21 Seconds
(without alignments)
942.928 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	179	7.5	1786	3	US-08-973-462-8	Sequence 8, Appli
2	172.5	7.2	1164	3	US-08-923-992A-2	Sequence 2, Appli
3	169	7.0	1098	3	US-08-923-992A-8	Sequence 8, Appli
4	168	7.0	1164	3	US-08-923-992A-10	Sequence 10, Appl
5	167.5	7.0	957	4	US-09-914-259-16	Sequence 16, Appl
6	166	6.9	1104	3	US-08-923-992A-4	Sequence 4, Appli
7	166	6.9	1128	3	US-08-923-992A-6	Sequence 6, Appli
8	166	6.9	2662	4	US-09-595-684B-31	Sequence 31, Appl
9	163.5	6.8	956	4	US-09-914-259-17	Sequence 17, Appl
10	162	6.8	8991	4	US-08-714-741-32	Sequence 32, Appl
11	159.5	6.6	816	2	US-08-533-306A-6	Sequence 6, Appli
12	159.5	6.6	816	2	US-08-742-923A-6	Sequence 6, Appli
13	158.5	6.6	885	2	US-08-533-306A-4	Sequence 4, Appli
14	158.5	6.6	885	2	US-08-742-923A-4	Sequence 4, Appli
15	156.5	6.5	1087	4	US-09-914-259-12	Sequence 12, Appl
16	156	6.5	1354	3	US-08-685-871-2	Sequence 2, Appli
17	154.5	6.4	1231	4	US-08-714-741-41	Sequence 41, Appl
18	153.5	6.4	984	1	US-08-242-932-2	Sequence 2, Appli
19	153.5	6.4	984	1	US-08-714-481-2	Sequence 2, Appli
20	153.5	6.4	984	5	PCT-US95-06111-2	Sequence 2, Appli
21	153	6.4	1588	5	PCT-US93-07261-11	Sequence 11, Appl
22	153	6.4	1663	5	PCT-US93-07261-16	Sequence 16, Appl
23	153	6.4	2285	4	US-09-308-375-2	Sequence 2, Appli
24	150.5	6.3	1010	4	US-09-134-001C-5178	Sequence 5178, Ap
25	147	6.1	534	4	US-09-103-664A-2	Sequence 2, Appli
26	146.5	6.1	1388	2	US-08-685-576-1	Sequence 1, Appli
27	146.5	6.1	1388	2	US-08-685-576-4	Sequence 4, Appli
28	145.5	6.1	652	4	US-08-559-896B-2	Sequence 2, Appli
29	145	6.0	588	4	US-08-714-741-42	Sequence 42, Appl
30	144.5	6.0	343	3	US-08-937-271-17	Sequence 17, Appl
31	144	6.0	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
32	143.5	6.0	864	4	US-08-714-741-40	Sequence 40, Appl
33	143	6.0	3878	4	US-09-914-259-11	Sequence 11, Appl
34	142	5.9	967	4	US-09-914-259-21	Sequence 21, Appl
35	142	5.9	1032	4	US-09-914-259-26	Sequence 26, Appl
36	141.5	5.9	1146	4	US-08-914-999-6	Sequence 6, Appli
37	141	5.9	1507	3	US-08-929-329-5	Sequence 5, Appli
38	139.5	5.8	963	4	US-09-914-259-20	Sequence 20, Appl
39	139	5.8	1939	4	US-09-310-187A-1	Sequence 1, Appli
40	139	5.8	3111	2	US-08-460-309-4	Sequence 4, Appli
41	139	5.8	3111	2	US-08-125-077-4	Sequence 4, Appli
42	138	5.8	630	3	US-08-973-462-9	Sequence 9, Appli
43	138	5.8	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
44	138	5.8	1234	4	US-09-592-054-8	Sequence 8, Appli
45	137.5	5.7	586	2	US-08-630-822A-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8

Query Match 7.5%; Score 179; DB 3; Length 1786;
Best Local Similarity 21.5%; Pred. No. 0.00012;
Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;

```
Qy          27 QDKSLHNRELSAE--RPLNEQIAEAEEDKIKKTYPPENKPGQSN-------YSFVDNL-- 75
               :: :: : |  ||    ||| :    ||:::|     :  ::|           ||:
Db         957 EENAVESNENVAENLEKLNETVFNTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKG 1016

Qy          76 ---NLLK----------AITEKEKIEKERQSIRSSPLDNKLN--------VEDVDSTK-- 112
                 |||            |  :||::   ::: || |||  |        :|:: ||:
Db        1017 IQENLLTGMFRSIETSIVIQSEEKVDL-NENVVSSILDNIENMKEGLLNKLENISSTEGV 1075

Qy         113 --------NRKLIDDYD-----------STKSGLDHKFQDDPDGLHQLDGTPLTAEDI- 151
                    : :  | |              :   ||   | :  | : : :   :| |:|
Db        1076 QETVTEHVEQNVYVDVDVPAMKDQFLGILNEAGGLKEMFFNLED-VFKSESDVITVEEIK 1134

Qy         152 -------VHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKE 204
                    | |    | || :  : |           : | :   | |:: : ::: |
Db        1135 DEPVQKEVEKETVSIIEEMEENIVD----------VLEEEKEDLTDKMIDAVEESIEIS 1183

Qy         205 ANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTK 264
               ::: ||        ||:    | :|   :: : : :   : ||:|   | | |  |  |
Db        1184 SDSKEE--------TES----IKDKEKDVSLVVEEVQDNDMDESVEKVLELKNMEEELMK 1231

Qy         265 TYSEDN------FEELQYFPNFYA-LLKSIDSEKEAK-----------EKETLITIMK 304
```

```
Db       1232 DAVEINDITSKLIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIIDAKDDTLEKVIE 1291

Qy        305 ------TLIDFVKMMVKYGTISPE--EGVSYLENLDEMIALQTK--NKLEKNATDNISKL 354

Db       1292 EEHDITTTLDEV---VELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESEILEDYKEL 1348

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRK 414

Db       1349 KTIETDILEEKKEIEKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLEEVHE-LKE 1407

Qy        415 NIEWLKKHDK--KGNKEDYDLSKMRDFINKQADAY---VEKGILDK----------EEA 458

Db       1408 EVEHIISGDAHIKGLEED-DLEEVDDLKGSILDMLKGDMELGDMDKESLEDVTTKLGERV 1466

Qy        459 EAIKRIYSS 467

Db       1467 ESLKDVLSS 1475


RESULT 2
US-08-923-992A-2
; Sequence 2, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1164 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-2

  Query Match             7.2%;  Score 172.5;  DB 3;  Length 1164;
  Best Local Similarity   21.6%;  Pred. No. 0.0002;
  Matches  124;  Conservative   77;  Mismatches  189;  Indels  185;  Gaps   25;

Qy         36 LSAERPLNE--------QIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT-EKEK 86

Db        149 LELENQFNETNRLLHIKQHEEVEKDKKAK----QQKTLKQSDTKVDLSNIDKELNHQKSQ 204

Qy         87 IEK--ERQSIRSSPLDNKL--------------NVEDVDSTKNRKLIDDYDSTKSGLDHK 130

Db        205 VEKMAEQKGITNEDKDSMLKKIEDIRKQAQQADKKEDAEVKVREELGKLFSSTKAGLDQE 264

Qy        131 FQDDPDGLHQLDGTPLTAEDIVHKI----------AARIYEENDRAVFDKIVSKLLNLG 179

Db        265 IQE-----HVKKET--SSEENTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQF 317

Qy        180 L------------------ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEE 210

Db        318 LENAQKLKEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEG 377

Qy        211 DPNKPTSWTE------------------NQAGKIPEKVTPMAAIQDGLAKGENDETVSN 251

Db        378 KLNQSKNLPELKQLEEEAHSKLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQ---Q 434

Qy        252 TLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVK 311

Db        435 KIELTVSPENITVYEGED--------VKFTVTAKS-DS--------------KTTLDFSD 471

Qy        312 MMVKYG-TISPEEGVSYLENLD----------------EMIALQTK----NKLEKNATD 349

Db        472 LLTKYNPSVSDRISTNYKTNTDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTI 531

Qy        350 NISKLFPAPSEKSHEETDSTKEEAA-----------------------KMEKEYGSLKD 385

Db        532 TVQKKEEKQVPKTPEQKDSKTEEKVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELME 591

Qy        386 STKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKK----HDKKGNKEDYDLSKMRDFIN 441

Db        592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMN 648

Qy        442 KQADAYVEKGILDKEEAE---------AIKRIYSS 467

Db        649 YQLHAQME--MLTRKVVQYMNKYPDNAEIKKIFES 681


RESULT 3
US-08-923-992A-8
; Sequence 8, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  8:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1098 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-8

  Query Match             7.0%;  Score 169;  DB 3;  Length 1098;
  Best Local Similarity   19.3%;  Pred. No. 0.00034;
  Matches  104;  Conservative   80;  Mismatches  182;  Indels  174;  Gaps   19;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLKAI 81
              | ::| |   : : :  ::  |   |:    ||   |     :       : | | |
Db         74 GKREKQLQQWKNNLKNDVDNTILSHEQKNEFKTKIDETNDSDALLELENQFNETNRLLHI 133

Qy         82 TEKEKIEKERQSIRSSPL---DNKLNVEDVDSTKNR----------------KLIDD-- 119
               : |::||:::: :   |   | |::: ::|   |                 | |:|
Db        134 KQHEEVEKDKKAKQQKTLKQSDTKVDLSNIDKELNHQKSQEAGITNEDKDSMLKKIEDIR 193

Qy        120 ----------------------YDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKI- 155
                                   : |||:||| : |:     |    |  ::|:   |:
Db        194 KQAQQPDKKEDAEVKVREELGKLFSSTKAGLDQEIQE-----HVKKET--SSEENTQKVD 246

Qy        156 ----------AARIYEENDRAVFDKIVSKLLNLGL-----------------ITESQA 186
                     | :  || |:|  ::  ::: | |                    : :: :
Db        247 EHYANSLQNLAQKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIKETNVKLYKAMS 306

Qy        187 HTLEDEVAEV-------LQKLISKE---ANNYEEDPNKPTSWTE---------------- 220
              :||    |:       |: |::|      ||   |: :  |
Db        307 ESLEQVEKELKHNSEANLEDLVAKSKEIVREYEGKLNQSKNLPELKQLEEEAHSKLKQVV 366

Qy        221 ---NQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYF 277
                 :  |  |:|||   ::  ||  ||::     : ||   |  |     ||
Db        367 EDFRKKFKTSEQVTPKKRLKRDLAANENNQ---QKIELTVSPENITVYEGED-------- 415

Qy        278 PNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYG-TISPEEGVSYLENLD---- 332
               |   || ||                || :||  :: ||  ::|     :|  | |
Db        416 VKFTVTAKS-DS-------------KTTLDFSDLLTKYNPSVSDRISTNYKTNTDNHKI 460

Qy        333 ------------EMIALQTK----NKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK 375
                         : : |: |    | :||  |  : |       |: |: ||  ||
Db        461 AEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTITVQKKEEKQVPKTPEQKDSKTEEKVP 520

Qy        376 MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSK 435
               | :             ||   : |     :  || : | |: | :  :  :  :| : |
Db        521 QEPK------------SNDKNQLQELIKSAQQELEKLEKAIKELMEQPEIPSNPEYGIQK 568


RESULT 4
US-08-923-992A-10
; Sequence 10, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1164 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-10

  Query Match             7.0%;  Score 168;  DB 3;  Length 1164;
  Best Local Similarity   19.0%;  Pred. No. 0.00044;
  Matches  121;  Conservative   85;  Mismatches  177;  Indels  254;  Gaps   25;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLKAI 81
              | ::| |   : : :  ::  |   |:    ||   |     :       : | | |
Db        105 GKREKQLQQWKNNLKNDVHNTILSHEQKNEFKTKIDETNDSDALLELENQFNETNRLLHI 164

Qy         82 TEKEKIEKERQSIRSSPL---DNKLNVEDVDSTKNR----------------------K 115
               : |::||:::: :   |   | |::: ::|   |                       |
Db        165 KQHEEVEKDKKAKQQKTLKQSDTKVDLSNIDKELNHQKSQVETMAEQLGITNEDKDSMLK 224

Qy        116 LIDD----------------------YDSTKSGLDHKFQDDPDGLHQLDGTPLTAED 150
               |:|                    : |||:||| : |:     |    |  |:|:
Db        225 KIEDIRKQAQQADKKEDAEVKVREELGKLFTSTKAGLDQEIQE-----HVKKET--TSEE 277

Qy        151 IVHKI-----------AARIYEENDRAVFDKIVSKLLN-----------LGLITE---- 183
                 |:           | :  || |:|  ::  ::: |           || |
Db        278 NTQKVDEHYPNSLQNLAQKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIKETNVK 337

Qy        184 ------------------SQAHTLEDEVAEVLQKLISKE-ANNYEEDPNKPTSWTE--- 220
                               |||: ||| ||:      |||    ||   |: :  |
Db        338 LYKAMSESLEQVEKQLKHNSQAN-LEDLVAK------SKEIVREYEGKLNQSKNLPELKQ 390

Qy        221 ----------------NQAGKIPEKVTPMAAIQDGLAKGENDE-----TVS-NTLTLTNG 258
                             :  |  |:|||   ::  ||  ||::     |||   :|:  |
Db        391 LEEEAHSKLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQQKIELTVSPENITVYEG 450

Qy        259 LERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYG- 317
                       |:|:     : |    ||              || :||  :: ||
Db        451 -------------EDLK-----FTLTAKSDS-------------KTTLDFSDLLTKYNP 478

Qy        318 TISPEEGVSYLENLD----------------EMIALQTK----NKLEKNATDNISKLFP 356
               ::|     :|  | |                : : |: |    | ::|  |  : |
Db        479 SVSDRISTNYKTNTDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVQKTFTITVQKKEE 538

Qy        357 APSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNI 416
                 |: |: ||  ||    | :              ||   :  |     :  || : | |
Db        539 KQVPKTPEQKDSKTEEKVPQEPK-----------SNDKNQLQELIKSAQQQLEKLEKAI 586

Qy        417 EWLKKHDKKGNKEDYDLSK----------------------------------MRDF 439
               : | :  :  :  :| : |                                  ||
Db        587 KELMEQPEIPSNPEYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDF 646

Qy        440 INKQADAYVEKGILDKEEAE---------AIKRIYSS 467
               :| |  | :|  :| ::  :          ||:|: |
Db        647 MNYQLHAQME--MLTRKVVQYINKYPDNAEIKKIFES 681


RESULT 5
US-09-914-259-16
; Sequence 16, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
```

```
;  APPLICANT: Makowski, Lee
;  APPLICANT:  Hyman, Paul
;  APPLICANT:  Williams, Mark
;  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
;  FILE REFERENCE: 8471-010-999
;  CURRENT APPLICATION NUMBER: US/09/914,259
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 180
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 957
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-914-259-16

  Query Match             7.0%;  Score 167.5;  DB 4;  Length 957;
  Best Local Similarity   20.3%;  Pred. No. 0.00036;
  Matches  103;  Conservative   98;  Mismatches  182;  Indels  125;  Gaps   23;

Qy         25 GSQDKSLHNR-----ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLL- 78
Db        321 GQRAKTIKNTVSVNLELTAE------------EWKKKYEKEKEKNKTLKNVIQHLEMEL 367

Qy         79 ------KAITEKEKIEKERQSIRSSPLDNKLNVEDVD------STKNRKLIDDYDSTKSG 126
Db        368 NRWRNGEAVPEDEQISAKDQK-NLEPCDNTPIIDNIAPVVAGISTEEK---EKYDEEISS 423

Qy        127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEND-----RAVFDKIVSKLLNLGLI 181
Db        424 LYRQLDDKDDEINQ-------QSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIE 476

Qy        182 TESQAHTLEDEVAEVLQKLISKEAANNYE---EDPNKPTSWTENQAGKIPEKVTPMAAIQD 238
Db        477 NEA----AKDEVKEVLQAL-EELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQR 531

Qy        239 GLAK-----------------------------GENDETVSNTLTLTNGLERR----TK 264
Db        532 ELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTND---VKTLADVNGVIEEEFTMAR 588

Qy        265 TYSEDNFEELQYFPNFYALLKS--IDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPE 322
Db        589 LYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLT---- 644

Qy        323 EGVSYLENLDEMIALQTKNKLEK---NATDNISKLFPAPSEKSHEETDSTKE-------- 371
Db        645 ---DYMQNME-----QKRRQLEESQDSLSEELAKL--RAQEKMHEVSFQDKEKEHLTRLQ 694

Qy        372 EAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDY 431
Db        695 DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEK-LSSDY 753

Qy        432 DLSKMRDFINKQADAYVEKGIL--DKEE 457
Db        754 NKLKIED---QEREMKLEKLLLLNDKRE 778


RESULT 6
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1104 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-4

  Query Match             6.9%;  Score 166;  DB 3;  Length 1104;
  Best Local Similarity   21.2%;  Pred. No. 0.00057;
  Matches  121;  Conservative   80;  Mismatches  190;  Indels  181;  Gaps   26;

Qy         17 IQAFPKPGGSQDKSLHNREL-----SAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF 71
Db        139 VEKHNKP--NQQKTLKQSDTKVDLSNIDKELNHQKSQVEA-----------MAEQAGITN 185

Qy         72 VDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKF 131
Db        186 EDKDSMLKKI---EDIRKQAQQ-----ADKK---EDAEVKVREELGKLFSSTKAGLDQQI 234

Qy        132 QDDPDGLHQLDGTPLTAEDIVHKI-----------AARIYEENDRAVFDKIVSKLLNLGL 180
Db        235 QE-----HVKKET--SSEENTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQFL 287

Qy        181 -------------------ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEED 211
Db        288 ENAQKLKEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLQDLVAKSKEIVREYEGK 347

Qy        212 PNKPTSWTE------------------NQAGKIPEKVTPMAAIQDGLAKGENDETVSNT 252
Db        348 LNQSKNLPELKQLEEEAHSKLKQVVEHFRKKFKTSEQVTPKKRVKRDLAANENNQ---QK 404

Qy        253 LTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKM 312
Db        405 IELTVSPENITVYEGED--------VKFTVTAKS-DS--------------KTTLDFSDL 441

Qy        313 MVKYG-TISPEEGVSYLENLD-----------------EMIALQTK----NKLEKNATDN 350
Db        442 LTKYNPSVSDRISTNYKTNTDNHKIAEITIKNLKLNQSQTVTLKAKDDSGNVVEKTFTIT 501

Qy        351 I----SKLFPAPSEKSHEETD-------------------STKEEAAKMEKEYGSLKDS 386
Db        502 VQKKEEKQVPKTPEQKHSKTEQNVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELMEQ 561

Qy        387 TKDDNSNP--GGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQA 444
Db        562 -PEIPSNPEYGIQKSIWESQKEPIQEAITSFNKIIGDSSSKYYTEHYFNKYKSHFMNYQL 620

Qy        445 DAYVEKGILDKEEAE---------AIKRIYSS 467
Db        621 HAQME--ILTRKVVQYMNKYPDNAEIKKIFES 650


RESULT 7
```

```
US-08-923-992A-6
; Sequence 6, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1128 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-6

  Query Match             6.9%;  Score 166;  DB 3;  Length 1128;
  Best Local Similarity   19.3%;  Pred. No. 0.00059;
  Matches  119;  Conservative   86;  Mismatches  200;  Indels  210;  Gaps   23;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSF---VDNLNLLKAI 81
Db         69 GKREKQLQQWKNNLKNDVDNTILSHEQKNEFKTKIDETNDSDALLELENQFNETNRLLHI 128

Qy         82 TEKEKIEKERQSIRSSPL---DNKLNVEDVDSTKNR----------------------K 115
Db        129 KQHEEVEKDKKAKQQKTLKQSDTKVDLSNIDKELNHQKSPVEKMAEPKGITNEDKDSMLK 188

Qy        116 LIDD------------------------YDSTKSGLDHKFQDDPDGLHQLDGTPLTAED 150
Db        189 KIEDIRKQAQQADKKEDAEVKVREELGKLFSSTKAGLDQE-------IHEHVKKETSSEE 241

Qy        151 IVHKI-----------AARIYEENDRAVFDKIVSKLLNLGL------------------ 180
Db        242 NTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQFLENAQKLKEMQPLIKETNVK 301

Qy        181 ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEEDPNKPTSWTE---------- 220
Db        302 LYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEGKLNQSKNLPELKQLEEEAHS 361

Qy        221 ---------NQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNF 271
Db        362 KLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQ---QKIELTVSPENITVYEGED-- 416

Qy        272 EELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYG-TISPEEGVSYLEN 330
```

```
Db        417 ------VKFTVTAKS-DS--------------KTTLDFSDLLTKYNPSVSDRISTNYKTN 455

Qy        331 LD-----------------EMIALQTK----NKLEKNATDNISKLFPAPSEKSHEETDST 369
Db        456 TDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTITVQKKEEKQVPKTPEQKDSK 515

Qy        370 KEEAA-----------------------KMEKEYGSLKDSTKDDNSNPGGKTDEPKGKT 405
Db        516 TEEKVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELMEQ-PEIPSNP--EYGIQKSIW 572

Qy        406 EAYLEAIRKNIEWLKK----HDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE-- 459
Db        573 ESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMNYQLHAQME--MLTRKVVQYM 630

Qy        460 -------AIKRIYSS 467
Db        631 NKYPDNAEIKKIFES 645


RESULT 8
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Ohashi, Cara
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  VAisberg, Eugeni
;  APPLICANT:  Wood, Kenneth
;  APPLICANT:  Yu, Ming
;  TITLE OF INVENTION: Human kinesins and methods of producing
;  TITLE OF INVENTION:  and purifying human kinesins
;  FILE REFERENCE: cytop036
;  CURRENT APPLICATION NUMBER: US/09/595,684B
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 09/295,612
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 2662
;   TYPE: PRT
;   ORGANISM: Human
US-09-595-684B-31

  Query Match             6.9%;  Score 166;  DB 4;  Length 2662;
  Best Local Similarity   21.8%;  Pred. No. 0.0019;
  Matches  119;  Conservative  103;  Mismatches  175;  Indels  150;  Gaps   28;

Qy         26 SQDKSLHNRELSAERPLN---EQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT 82
Db       1101 AQEK---NHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQL-------LNVQEEMS 1150

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTK---NRKLIDDYDSTKS---------GLDHK 130
Db       1151 EMQKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVLKELQKS 1210

Qy        131 FQDDPDGL----HQLDGTPL-TAEDI----VH-KIAARIYEENDRAVFDKIVSKLLNLGL 180
Db       1211 FETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEK-TAQIINTQD 1269

Qy        181 ITESQAHTLEDEVAEVL---QKLIS--KEANNYEEDPNKPTSWTENQAGKIPEKVTPMAA 235
Db       1270 LEKS--HTKLQEEIPVLHEEQELLPNVKKVSETQETMNELELLTEQSTTK---DSTTLAR 1324

Qy        236 IQ-----------------DGLAK-GENDETVSNTL-------------TLTNGLERRTK 264
Db       1325 IEMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSK 1384

Qy        265 TYSEDNFEEL------------QYFPNFYALLK--------------SIDSEKE-AKEKE 297
Db       1385 QEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKD 1444
```

```
Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIA--LQTKNKLE------KNATD 349
Db       1445 DLQRLQEVL-------------QSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEE 1490

Qy        350 NISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTE--- 406
Db       1491 TINELRVNLSEK--------ETEISTIQKQLEAINDKLQN-------KIQEIYEKEEQLN 1535

Qy        407 -AYLEAIRKNIEWLKKHDKKGNKEDYDL----SKMRDFINKQADAYVEKGILDKEEAEAI 461
Db       1536 IKQISEVQENVNELKQFKEHRKAKDSALQSIESKMLELTNRLQESQEEIQIMIKEK-EEM 1594

Qy        462 KRIYSSL 468
Db       1595 KRVQEAL 1601


RESULT 9
US-09-914-259-17
; Sequence 17, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
;  APPLICANT: Makowski, Lee
;  APPLICANT:  Hyman, Paul
;  APPLICANT:  Williams, Mark
;  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
;  FILE REFERENCE: 8471-010-999
;  CURRENT APPLICATION NUMBER: US/09/914,259
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 180
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 956
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-914-259-17

  Query Match             6.8%;  Score 163.5;  DB 4;  Length 956;
  Best Local Similarity   20.7%;  Pred. No. 0.00072;
  Matches  106;  Conservative   91;  Mismatches  169;  Indels  147;  Gaps   25;

Qy         25 GSQDKSLHNR-----ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLL- 78
Db        321 GQRAKTIKNTVSVNLELTAE------------EWKKKYEKEKEKNKALKSVLQHLEMEL 367

Qy         79 ------KAITEKEKIE-KERQSIRSSPLDNKLNVED----VDSTKNRKLIDDYDSTKSGL 127
Db        368 NRWRNGEAVPEDEQISAKDHKSL--EPCDNTPIIDNITPVVDGISAEK--EKYDEEITSL 423

Qy        128 DHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEND-----RAVFDKIVSKLLNLGLIT 182
Db        424 YRQLDDKDDEINQ-------QSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIEN 476

Qy        183 ESQAHTLEDEVAEVLQKLISKEANNYE---EDPNKPTSWTENQAGKIPEKVTPMAAIQDG 239
Db        477 EA----AKDEVKEVLQAL-EELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRE 531

Qy        240 LAK-----------------------------GENDETVSNTLTLTNGLERR----TKT 265
Db        532 LSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTND---VKTLADVNGVIEEEFTMARL 588

Qy        266 YSEDNFEELQYFPNFYALLKS--IDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEE 323
Db        589 YISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLT----- 643

Qy        324 GVSYLENLDEMIALQTKNKLEK---NATDNISKLFPAPSEKSHEET--DSTKEEAAKMEK 378
Db        644 --DYMQNME-----QKRRQLEESQDSLSEELAKL--RAQEKMHEVSFQDKEKEHLTRLQ- 693

Qy        379 EYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRD 438
Db        694 -----------------DAEEVKKALEQQMESHR------EAHQKQ-------LSRLRD 722


Qy        439 FINKQADAYVEKGILDKE---EAEAIKRIYSSL 468
Db        723 EIEEKQRIIDEIRDLNQKLQLEQERLSSDYNKL 755


RESULT 10
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  McDaniel, Larry S.
;    APPLICANT:  Swiatlo, Edwin
;    APPLICANT:  Yother, Janet
;    APPLICANT:  Crain, Marilyn J.
;    APPLICANT:  Hollingshead, Susan
;    APPLICANT:  Tart, Rebecca
;    APPLICANT:  Brooks-Walter, Alexis
;    TITLE OF INVENTION:  PNEUMOCOCCAL GENES, PORTIONS THEREOF,
;    TITLE OF INVENTION:  EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
;    TITLE OF INVENTION:  PORTIONS AND PRODUCTS
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,741
;      FILING DATE:  16-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer Esq., William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454312-2460
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8991 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  amino acid
US-08-714-741-32

  Query Match             6.8%;  Score 162;  DB 4;  Length 8991;
  Best Local Similarity   18.2%;  Pred. No. 0.02;
  Matches   88;  Conservative  110;  Mismatches  184;  Indels  102;  Gaps   19;

Qy         26 SQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEKE 85
Db       2024 AQEKYADYQRRIEEKAAKEQQASLEQQEANKDYQLKLKKYLDGRNLSNSSVLKKEMEEAE 2083

Qy         86 KIEKERQSIRSSPLDNKL-------NVEDVDSTKNRKLIDDYDSTKSGLDHKFQD----- 133
Db       2084 KKDKEKQA-----EFNKIRREIVVPNPQELEMARRKSEV--VKAKESGLVKRVEEAEKKV 2136

Qy        134 -------DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS-KLLNLGLITESQ 185
Db       2137 TEARQKLDAERAKEVVLQPTRVENEVHKLXQKMNKK--KMILTSLASVAILGAGLVTSQP 2194

Qy        186 AHTLEDEVAEVLQK----------------------LISKEANN----YEEDPNKPTSWT 219
Db       2195 TFVRAEESPQVVEKSSLEKKYEEAKAKADTAKKDYETAKKKAEDAQKKYEDDQKR----T 2250
```

```
Qy        220 ENQAGKIPE---KVTPMA----------AIQDGLAKGENDETVSNTLT-LTNGLERRTK 264
Db       2251 EEKARKEAEASQKLIDVALVVQNAYKEYREVQNQRSKYKSDADYQKKLTEVDSKIEKARK 2310

Qy        265 TYS--EDNFEELQYFPNFYALL----KSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGT 318
Db       2311 EQQDLQNNFNEVR------AVVAPDPTCVGXDXRMNKKKMILTSLASV-----AILGAGX 2359

Qy        319 ISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAKMEK 378
Db       2360 VTSQPTXVRAEEAPQVV---EKSSLEKKYEEAKAKY-----DAAKKDYDEAKKKAAEAQK 2411

Qy        379 EYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGN---KEDYDLSK 435
Db       2412 KYEEDQKKTEEKAEKAKAASEEIAKATEEVQKAVLDYITAIRNHNDSGKTSAEEAENKAK 2471

Qy        436 MRDF 439
Db       2472 ERDY 2475
```

RESULT 11
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 6.6%; Score 159.5; DB 2; Length 816;
Best Local Similarity 22.6%; Pred. No. 0.0011;
Matches 109; Conservative 81; Mismatches 193; Indels 99; Gaps 21;

```
Qy         26 SQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITE 83
Db        158 SHREEMENEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTK 215

Qy         84 KEKIEKERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KF 131
Db        216 LRQLEEERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRF 275

Qy        132 QDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQA 186
Db        276 QKEIENLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQ 328

Qy        187 HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGEN 245
Db        329 RKFDQLLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEE 379

Qy        246 DETVSNTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLIT 301
Db        380 LERTNKMLK----AEMEDLVSSKDDVGKNVHEL---------------EKSKRALETQME 420

Qy        302 IMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------L 354
Db        421 EMKTQLEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQL 476

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAI 412
Db        477 HEYETELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDF 535

Qy        413 RKNIE-----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDK 455
Db        536 QRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEK 588

Qy        456 EE 457
Db        589 EE 590
```

RESULT 12
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: No. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```
;    MOLECULE TYPE:  protein
US-08-742-923A-6

  Query Match             6.6%;  Score 159.5;  DB 2;  Length 816;
  Best Local Similarity   22.6%;  Pred. No. 0.0011;
  Matches  109;  Conservative   81;  Mismatches  193;  Indels   99;  Gaps   21;

Qy         26 SQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITE 83
              |  : : |    |    |||  || :  |: |      :   |      :  :    |   |:
Db        158 SHREEMENEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTK 215

Qy         84 KEKIEKERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KF 131
               ::|:|| |::      :: | |:|   || |      :| : |: ||   |:    :|
Db        216 LRQLEEERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRF 275

Qy        132 QDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQA 186
              | : : | |        | | : |   |: :| |  | |      ::|| |        | :
Db        276 QKEIENLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQ 328

Qy        187 HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGEN 245
              :  :||  :| ||   : | :: ::     | :| :   |   :| |:::  |   |
Db        329 RKFDQLLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEE 379

Qy        246 DETVSNTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLIT 301
              |  :  |      |      |:|     |  ||              ||  :  || :
Db        380 LERTNKMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQME 420

Qy        302 IMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------L 354
              |||  :: ::  ::    : |:    ||   : :  | :  |:      |  |        |
Db        421 EMKTQLEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQL 476

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAI 412
              :|   |  :     |||  : | | |||     :|   |: :  |   | :| ::
Db        477 HEYETELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDF 535

Qy        413 RKNIE-----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDK 455
              :: :|           |:::||    : || :::: :       ||||       |:|
Db        536 QRELEDARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEK 588

Qy        456 EE 457
              ||
Db        589 EE 590


RESULT 13
US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869COB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-533-306A-4

  Query Match             6.6%;  Score 158.5;  DB 2;  Length 885;
  Best Local Similarity   22.9%;  Pred. No. 0.0015;
  Matches  109;  Conservative   79;  Mismatches  190;  Indels   99;  Gaps   21;

Qy         31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITEKEKIE 88
              | |    |    |||  || :  |: |      :   |      :  :    |   |:  ::|
Db        232 LQNEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTKLRQLE 289

Qy         89 KERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KFQDDPD 136
              :|| |::      :: | |:|   || |      :| : |: ||   |:    :|| : :
Db        290 EERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRFQKEIE 349

Qy        137 GLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQAHTLED 191
              | |        | | : |   |: :| |  | |      ::|| |        | :    :
Db        350 NLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQRKFDQ 402

Qy        192 EVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGENDETVS 250
              :||  :| ||   : | :: ::     | :| :   |   :| |::: |   |  | :
Db        403 LLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEELERTN 453

Qy        251 NTLTLTNGLERRTKTYSED----NFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTL 306
              |      |      |:|     |  ||              ||  :  || :  |||
Db        454 KMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQMEEMKTQ 494

Qy        307 IDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------LFPAPS 359
              :: ::  ::    : |:    ||   : :  | :  |:      |  |       |    :
Db        495 LEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYET 550

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAIRKNIE 417
              |   |  :     |||  : | | |||     :|   |: :  |   | :| ::  :: :|
Db        551 ELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELE 609

Qy        418 -----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDKEE 457
                        |:::||    : || :::: :       ||||       |:|||
Db        610 DARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEKEE 659


RESULT 14
US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,923A
;      FILING DATE:  No. 5869611ember 1, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869DVC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-742-923A-4

  Query Match             6.6%;  Score 158.5;  DB 2;  Length 885;
  Best Local Similarity   22.9%;  Pred. No. 0.0015;
  Matches  109;  Conservative   79;  Mismatches  190;  Indels   99;  Gaps   21;

Qy         31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFV--DNLNLLKAITEKEKIE 88
Db        232 LQNEVESVTGMLNE--AEGKAIKLAKDVASLSSQLQDTQELLQEETRQKLNVSTKLRQLE 289

Qy         89 KERQSIR---SSPLDNKLNVEDVDSTKN------RKLIDDYDSTKSGLDH---KFQDDPD 136
Db        290 EERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKKLQDFASTVEALEEGKKRFQKEIE 349

Qy        137 GLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD-----KIVSKLLNLGLITESQAHTLED 191
Db        350 NLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQRQLVSNL-------EKKQRKFDQ 402

Qy        192 EVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMA-AIQDGLAKGENDETVS 250
Db        403 LLAE--EKNIS---SKYADERDR----AEAEAREKETKALSLARALEEALEAKEELERTN 453

Qy        251 NTLTLTNGLERRTKTYSED----NPEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTL 306
Db        454 KMLK----AEMEDLVSSKDDVGKNVHEL--------------EKSKRALETQMEEMKTQ 494

Qy        307 IDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISK-------LFPAPS 359
Db        495 LEELEDELQ----ASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYET 550

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPK--GKTEAYLEAIRKNIE 417
Db        551 ELEDERNERALAAAAKKKLE-GDLKDLELQADSAIKGREEAIKQLRKLQAQMKDFQRELE 609

Qy        418 -----------WLKKHDKKGNKEDYDLSKMRDFI------NKQADAYVEKGILDKEE 457
Db        610 DARASRDEIFATAKENEKKAKSLEADLMQLQEDLAAAERARKQAD-------LEKEE 659


RESULT 15
US-09-914-259-12
; Sequence 12, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
;  APPLICANT: Makowski, Lee
;  APPLICANT:  Hyman, Paul
;  APPLICANT:  Williams, Mark
;  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
;  FILE REFERENCE: 8471-010-999
;  CURRENT APPLICATION NUMBER: US/09/914,259
;  CURRENT FILING DATE:  2000-11-21
;  NUMBER OF SEQ ID NOS: 180
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;  LENGTH: 1087
;  TYPE: PRT
;  ORGANISM: Oryctolagus cuniculus
US-09-914-259-12

  Query Match             6.5%;  Score 156.5;  DB 4;  Length 1087;
  Best Local Similarity   19.3%;  Pred. No. 0.0028;
  Matches  103;  Conservative   83;  Mismatches  180;  Indels  167;  Gaps   22;

Qy         31 LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEKEKIEKE 90
Db        304 IHEKEVEIDR-LNEQIIKLQQQ-----------------------LKITTDNKVIEEK 337

Qy         91 RQSIRSSPLDNKLNVEDVDST-KNR--------KLIDDYDSTKSGLDHKFQDDPDGLHQL 141
Db        338 NELIRDLEAQIECLMSDQERVRKNREEEIEQLNEVIEKLQQELANIDQKTSVDPSSLSE- 396

Qy        142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVA------- 194
Db        397 -----EADSLKHQL------------DKVIAEKLAL----EHQVETTNEEMAVTKNVLK 434

Qy        195 -------EVLQKLIS--KEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAI--QDGLAKG 243
Db        435 ETNFKMNQLTQELCSLKRERERKMERIQSVPEKSVNMSVGDLSKDKPEMDLIPTEDALAQL 494

Qy        244 E-------NDETVSNTL-TLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKE 295
Db        495 ETQTQLRSSEESSKVSLSSLETKLLQLESTVSTKDLELTQCY-------KQIQDMREQGR 547

Qy        296 KET------LITIMKTL-----------------IDFVKMMVKYGTISPEEG-------- 324
Db        548 SETEMLQTKIVSLQKVLEEKVAAALVSQVQLEAVQEYVKLCADKPAVSSDPARTEVPGLS 607

Qy        325 -----------------VSYLENL-----------DEMIALQTKNKLEK-NATDNISKLF 355
Db        608 QLAGNTMESDVSALTWRISELESQLVEMHSSLISEKEQVEIAEKNALEKEKKLQELQKLV 667

Qy        356 PAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPG------GKTDEPKGKTEAYL 409
Db        668 QDSETKQRE-----RERQSRLHGDLGVLESTTSEESGVFGELEALRAESAAPKGELANYK 722

Qy        410 EAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADA---YVEKGILDKEEAE 459
Db        723 ELAEKLQEELLVKETNMASLPKELSHVRDQLTEAEDKLSHFSEKE--DKTEVQ 773


Search completed: January  2, 2004, 11:03:01
Job time : 23 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 11:00:07 ; Search time 38 Seconds
(without alignments)
2471.055 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2399	100.0	468	11	US-09-931-836-67	Sequence 67, Appl
2	2399	100.0	468	12	US-10-035-977-67	Sequence 67, Appl
3	2399	100.0	468	12	US-10-063-735-150	Sequence 150, App
4	2399	100.0	468	12	US-10-137-870-90	Sequence 90, Appl
5	2399	100.0	468	12	US-10-140-018-90	Sequence 90, Appl
6	2399	100.0	468	12	US-10-140-021-90	Sequence 90, Appl
7	2399	100.0	468	12	US-10-140-274-90	Sequence 90, Appl
8	2399	100.0	468	12	US-10-140-471-90	Sequence 90, Appl
9	2399	100.0	468	12	US-10-140-807-90	Sequence 90, Appl
10	2399	100.0	468	12	US-10-140-922-90	Sequence 90, Appl
11	2399	100.0	468	12	US-10-140-924-90	Sequence 90, Appl
12	2399	100.0	468	12	US-10-140-926-90	Sequence 90, Appl
13	2399	100.0	468	12	US-10-141-698-90	Sequence 90, Appl
14	2399	100.0	468	12	US-10-141-702-90	Sequence 90, Appl
15	2399	100.0	468	12	US-10-141-704-90	Sequence 90, Appl
16	2399	100.0	468	12	US-10-142-421-90	Sequence 90, Appl
17	2399	100.0	468	12	US-10-142-432-90	Sequence 90, Appl
18	2399	100.0	468	12	US-10-142-767-90	Sequence 90, Appl
19	2399	100.0	468	12	US-10-143-033-90	Sequence 90, Appl
20	2399	100.0	468	12	US-10-144-994-90	Sequence 90, Appl
21	2399	100.0	468	12	US-10-145-628-90	Sequence 90, Appl
22	2399	100.0	468	12	US-10-145-631-90	Sequence 90, Appl
23	2399	100.0	468	12	US-10-145-633-90	Sequence 90, Appl
24	2399	100.0	468	12	US-10-145-746-90	Sequence 90, Appl
25	2399	100.0	468	12	US-10-145-748-90	Sequence 90, Appl
26	2399	100.0	468	12	US-10-145-823-90	Sequence 90, Appl
27	2399	100.0	468	12	US-10-145-826-90	Sequence 90, Appl
28	2399	100.0	468	12	US-10-145-870-90	Sequence 90, Appl
29	2399	100.0	468	12	US-10-145-876-90	Sequence 90, Appl
30	2399	100.0	468	12	US-10-145-959-90	Sequence 90, Appl
31	2399	100.0	468	12	US-10-146-724-90	Sequence 90, Appl
32	2399	100.0	468	12	US-10-146-725-90	Sequence 90, Appl
33	2399	100.0	468	12	US-10-146-795-90	Sequence 90, Appl
34	2399	100.0	468	12	US-10-147-495-90	Sequence 90, Appl
35	2399	100.0	468	12	US-10-147-501-90	Sequence 90, Appl
36	2399	100.0	468	12	US-10-147-504-90	Sequence 90, Appl
37	2399	100.0	468	12	US-10-147-506-90	Sequence 90, Appl
38	2399	100.0	468	12	US-10-147-509-90	Sequence 90, Appl
39	2399	100.0	468	12	US-10-147-510-90	Sequence 90, Appl
40	2399	100.0	468	12	US-10-147-511-90	Sequence 90, Appl
41	2399	100.0	468	12	US-10-147-529-90	Sequence 90, Appl
42	2399	100.0	468	12	US-10-152-397-90	Sequence 90, Appl
43	2399	100.0	468	12	US-10-153-586-90	Sequence 90, Appl
44	2399	100.0	468	12	US-10-158-783-90	Sequence 90, Appl
45	2399	100.0	468	12	US-10-158-786-90	Sequence 90, Appl

ALIGNMENTS

RESULT 1
US-09-931-836-67
; Sequence 67, Application US/09931836
; Publication No. US20030027249A1
; GENERAL INFORMATION:
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3030R1C1
; CURRENT APPLICATION NUMBER: US/09/931,836
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/112514
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114140
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115552
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22

; PRIOR APPLICATION NUMBER: 60/125774
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125778
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125826
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/127035
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/127706
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/130359
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131272
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/132371
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132379
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132383
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/135750
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/138166
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/146970
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/162506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/644848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 09/747259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/816744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/854208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/854280
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 67
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-931-836-67

Query Match 100.0%; Score 2399; DB 11; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 2
US-10-035-977-67
; Sequence 67, Application US/10035977
; Publication No. US20030134327A1
; GENERAL INFORMATION:
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.

; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3030R1C10
; CURRENT APPLICATION NUMBER: US/10/035,977
; CURRENT FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/112514
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114140
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115552
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/125774
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125778
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125826
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/127035
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/127706
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/130359
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131272
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/132371
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132379
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132383
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/135750
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/138166
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/146970
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/162506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/644848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 09/747259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/816744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/854208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/854280
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 67
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-035-977-67

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
```

```
Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 3
US-10-063-735-150
; Sequence 150, Application US/10063735
; Publication No. US20030138882A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,735
;  CURRENT FILING DATE:  2002-05-08
;      Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 150
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-735-150

  Query Match            100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 4
US-10-137-870-90
; Sequence 90, Application US/10137870
; Publication No. US20030138883A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C155
;  CURRENT APPLICATION NUMBER: US/10/137,870
;  CURRENT FILING DATE:  2002-05-03
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-137-870-90

  Query Match            100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 5
US-10-140-018-90
; Sequence 90, Application US/10140018
; Publication No. US20030138885A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C158
; CURRENT APPLICATION NUMBER: US/10/140,018
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-018-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 6

US-10-140-021-90
; Sequence 90, Application US/10140021
; Publication No. US20030138886A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C167
; CURRENT APPLICATION NUMBER: US/10/140,021
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-021-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 7
US-10-140-274-90

; Sequence 90, Application US/10140274
; Publication No. US20030143674A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C161
; CURRENT APPLICATION NUMBER: US/10/140,274
; CURRENT FILING DATE: 2002-05-06
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-274-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 8
US-10-140-471-90
; Sequence 90, Application US/10140471
; Publication No. US20030138887A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C163
; CURRENT APPLICATION NUMBER: US/10/140,471
; CURRENT FILING DATE: 2002-05-06
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEAANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 9
US-10-140-807-90
; Sequence 90, Application US/10140807
; Publication No. US20030134354A1

```
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C174
;  CURRENT APPLICATION NUMBER: US/10/140,807
;  CURRENT FILING DATE:  2002-05-07
;  Prior Apploication removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-807-90

  Query Match             100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 10
US-10-140-922-90
; Sequence 90, Application US/10140922
; Publication No. US20030138889A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C179
;  CURRENT APPLICATION NUMBER: US/10/140,922
;  CURRENT FILING DATE:  2002-05-07
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-922-90

  Query Match             100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 11
US-10-140-924-90
; Sequence 90, Application US/10140924
; Publication No. US20030134355A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
```

; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C177
; CURRENT APPLICATION NUMBER: US/10/140,924
; CURRENT FILING DATE: 2002-05-07
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-924-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 12
US-10-140-926-90
; Sequence 90, Application US/10140926
; Publication No. US20030134356A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen

; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C187
; CURRENT APPLICATION NUMBER: US/10/140,926
; CURRENT FILING DATE: 2002-05-07
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-926-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 13
US-10-141-698-90
; Sequence 90, Application US/10141698
; Publication No. US20030134357A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura

; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C206
; CURRENT APPLICATION NUMBER: US/10/141,698
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-698-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 14
US-10-141-702-90
; Sequence 90, Application US/10141702
; Publication No. US20030134358A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C208
; CURRENT APPLICATION NUMBER: US/10/141,702
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-702-90

Query Match 100.0%; Score 2399; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 8.7e-150;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
```

RESULT 15
US-10-141-704-90
; Sequence 90, Application US/10141704
; Publication No. US20030134359A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen

```
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C209
;  CURRENT APPLICATION NUMBER: US/10/141,704
;  CURRENT FILING DATE:  2002-05-08
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 90
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-141-704-90

  Query Match           100.0%;  Score 2399;  DB 12;  Length 468;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-150;
  Matches  468;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


Search completed: January  2, 2004, 11:03:51
Job time : 39 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 21 Seconds
(without alignments)
2143.188 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1957	81.6	533	2	A37180	chromogranin/secre
2	196	8.2	2139	2	T18296	myosin heavy chain
3	191	8.0	2464	1	QRMSP1	microtubule-associ
4	186.5	7.8	2116	2	A26655	myosin heavy chain
5	184	7.7	665	2	B71609	hypothetical prote
6	183.5	7.6	1790	2	S67593	transport protein
7	180.5	7.5	2364	2	A56577	microtubule-associ
8	179.5	7.5	1526	2	A45605	mature-parasite-in
9	179	7.5	853	2	T51505	hypothetical prote
10	179	7.5	1005	2	A64465	hypothetical prote
11	179	7.5	1558	2	B71603	RESA-H3 antigen PF
12	179	7.5	2269	2	T28677	rhoptry protein -
13	178.5	7.4	1392	2	A43336	microtubule-vesicl
14	178.5	7.4	1427	2	S22695	restin - human
15	177	7.4	1871	2	D96796	probable heat shoc
16	172.5	7.2	1134	2	A60234	IgA Fc receptor pr
17	172.5	7.2	1164	1	FCSOAG	IgA Fc receptor pr
18	172.5	7.2	1979	2	C71622	hypothetical prote
19	171.5	7.1	3488	2	T34418	hypothetical prote
20	170	7.1	1804	2	T34518	nestin - golden ha
21	169	7.0	1272	2	C90593	hypothetical prote
22	168.5	7.0	746	2	T47237	myosin II heavy ch
23	168.5	7.0	926	2	T24923	hypothetical prote
24	168	7.0	2401	2	T28676	rhoptry protein -
25	167	7.0	976	2	T01553	hypothetical prote
26	166.5	6.9	1875	2	S38173	myosin-like protei
27	166.5	6.9	5327	2	T13564	microtubule-associ
28	166	6.9	991	2	H86168	hypothetical prote
29	166	6.9	1192	2	A71623	probable secreted
30	166	6.9	2663	1	S28261	centromere protein
31	164.5	6.9	719	2	A81358	hypothetical prote
32	164.5	6.9	911	2	S51441	hypothetical prote
33	164.5	6.9	1127	2	T28317	ORF MSV156 hypothe
34	164.5	6.9	1269	2	F84730	probable myosin he
35	164	6.8	1909	2	A45592	liver stage antige
36	163	6.8	1252	2	B42771	reticulocyte-bindi
37	163	6.8	2510	2	T28160	hypothetical prote
38	162.5	6.8	1738	2	T14867	interaptin - slime
39	162.5	6.8	2385	2	A32491	myosin heavy chain
40	162.5	6.8	2411	2	B32491	myosin heavy chain
41	161.5	6.7	1302	1	JC6009	surface-located me
42	161.5	6.7	1992	1	S02771	myosin heavy chain
43	161	6.7	852	2	D72230	conserved hypothet
44	160	6.7	1354	2	S74244	serine/threonine-s
45	159.5	6.6	821	2	S67087	hypothetical prote

ALIGNMENTS

RESULT 1
A37180
chromogranin/secretogranin-like vesicle protein precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 30-Sep-1993
C;Accession: A37180
R;Ottiger, H.P.; Battenberg, E.F.; Tsou, A.P.; Bloom, F.E.; Sutcliffe, J.G.
J. Neurosci. 10, 3135-3147, 1990
A;Title: 1B1075: a brain- and pituitary-specific mRNA that encodes a novel chromogranin/s
A;Reference number: A37180; MUID:90376160; PMID:2204688
A;Accession: A37180
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-533 <OTT>

Query Match 81.6%; Score 1957; DB 2; Length 533;
Best Local Similarity 86.7%; Pred. No. 3.5e-87;
Matches 383; Conservative 21; Mismatches 38; Indels 0; Gaps 0;

```
Qy         16 PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL 75
              |  :|||| |||||||||||||||||||||||||| |||||||| |:|| : |:| ||||
Db         63 PNSSFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKTYPSESKPSERNFSSVDNL 122

Qy         76 NLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDP 135
              |||||||||| :|| :||||||| ||:|||:| |||||||| |:|||||||| | ||||
Db        123 NLLKAITEKETVEKAKQSIRSSPFDNRLNVDDADSTKNRKLTDEYDSTKSGLDRKVQDDP 182

Qy        136 DGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAE 195
              |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||
Db        183 DGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLNLGLITESQAHTLEDEVAE 242

Qy        196 VLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTL 255
              |||||||||||||| | |||| |||| |||||||||:|| |||    |||:||||||||
Db        243 ALQKLISKEANNYEEAPEKPTSRTENQDGKIPEKVTPVAATQDGFTNRENDDTVSNTLTL 302

Qy        256 TNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVK 315
              :|||||||  : :|:|||||||||||||| ||||||||||||||||||||||||||||||
Db        303 SNGLERRTNPHRDDDFEELQYFPNFYALLTSIDSEKEAKEKETLITIMKTLIDFVKMMVK 362

Qy        316 YGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK 375
              |||||||||||||||||| |||||||||| ||: ||||||| ||||||||||||||||||
Db        363 YGTISPEEGVSYLENLDETIALQTKNKLEKNTTDSKSKLFPAPPEKSHEETDSTKEEAAK 422

Qy        376 MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSK 435
              |||||||||||||||| |||||| |||||||||||||||||||:||||||||||||||||
Db        423 MEKEYGSLKDSTKDDNSNLGGKTDEAKGKTEAYLEAIRKNIEWLKKHNKKGNKEDYDLSK 482

Qy        436 MRDFINKQADAYVEKGILDKEE 457
              |||||:|||||||||||| :::
Db        483 MRDFINQQADAYVEKGILIRKK 504
```

RESULT 2
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C;Genetics:
A;Gene: mhcA
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 8.2%; Score 196; DB 2; Length 2139;
Best Local Similarity 21.0%; Pred. No. 0.1;
Matches 107; Conservative 108; Mismatches 191; Indels 104; Gaps 23;

```
Qy      20 FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPEN-------KPGQSNYSFV 72
           | : |  :|| : :       | :::||    :|||    ||       | |:
Db     859 FEEEGKKKDKEIED--------LKKKLAE----EIKKREAAENALASATAKTGELEAKIQ 906

Qy      73 D---------------------NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVD- 109
           |                    || :    :||::::   :::    |:||  ||::
Db     907 DLEDKISELESKLSAAELDKQELNLKIENLEEDKEELKETIDNLKGDLKDSKLKGEDLEV 966

Qy     110 ---------STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIY 160
                   :| |   :|  | | : :    :  | : :| |      |:    :
Db     967 EITELNSQINTLN-ATVDKKDKTIAEMQESIDEKEDEITKLKGDIKLLEEEKDDL----- 1020

Qy     161 EENDR----AVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPT 216
           | ||     |  | |  ||  : :    :    :||:|:: |:|  :|  |  :|
Db    1021 -EQDRADVSATKDDIAKKLNKITI----ECEDAKDEIAKLEQELEDEE--NKNKDLTNEL 1073

Qy     217 SWTENQAGKIPEKVTPMAAI-------QDGLAKG-ENDETVSNTLTLTNG-LERRTKTYS 267
           |: : |:  : :    |          :| |::  ||::  :  || |   ||::
Db    1074 QQTQLKLGETEKSLAAQVAATKKASDERDTLSQNLENEKLTTKNLTKTKADLEKKISGLK 1133

Qy     268 EDNFEELQYFPNFYALLKSID-----SEKEAKEKETLITIMKTLIDFV-KMMVKYGTISP 321
           :| :|:|:   |        |:     :::: || :  ||    :  :: |  :|
Db    1134 QD-YEDLEDDKN------KIEGDLRNAQRKIKELDDEITKGADVSQYLQKQKEEY----- 1181

Qy     322 EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK----ME 377
           |  :: ::   | |    ||| ||   :   :|     ::  :||:  ||:| |    :|
Db    1182 ESQIAKMQEEKEAIGNDVKNK-EKTIKE--KELEIQSLQEKLDETEVEKEDAEKKKKEIE 1238

Qy     378 KEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL-SKM 436
           ||  :|::  |::  :    |:: | | |  |:  :| :: :   ::|   :  || :::
Db    1239 KEMKALQEE-KENVESSKNSTEKDKKKLEDNLKDTQKKLDDMTADNEKLKAKAKDLEAQL 1297

Qy     437 RDFINKQADAYVEKGILDKEEAEAIKRIYS 466
           :  :    |  :  :|:|::|:: | : |
Db    1298 NEVQDNHEKAVADAELLNKKKAQSDKELNS 1327
```

RESULT 3
QRMSP1
microtubule-associated protein MAP1B - mouse
N;Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
C;Accession: S07549; S44387; A33645
R;Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A;Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A;Reference number: A33645; MUID:90094539; PMID:2480963
A;Accession: S07549
A;Molecule type: mRNA
A;Residues: 1-2464 <NOB>
A;Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000
R;Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A;Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A;Reference number: S44387; MUID:94234720; PMID:8179328
A;Accession: S44387
A;Status: preliminary
A;Molecule type: protein
A;Residues: 653-663,'IC' <SAN>
C;Superfamily: microtubule-associated protein MAP1B
C;Keywords: microtubule binding; phosphoprotein; tandem repeat
F;589-786/Domain: microtubule binding #status experimental <MTB>
F;589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-69
R-K-E/D-X)
F;1861-2064/Region: 17-residue repeats
F;91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: phc
F;147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (co
F;1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 8.0%; Score 191; DB 1; Length 2464;
Best Local Similarity 21.3%; Pred. No. 0.21;
Matches 118; Conservative 85; Mismatches 186; Indels 166; Gaps 24;

```
Qy      22 KPGGSQDKSLHNRELSAERPLNEQIAEA------EEDKIKKTYP--PENKPGQSNYSFVD 73
           ||  |:     ::| : |      |: :       |:  :||  |   |:||
Db    550 KPVASKSVRKESKEETPEVTKTSQVEKTPKVESKEKVLVKKDKPVKTESKP--------- 600

Qy      74 NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQD 133
             ::||||   || ||   : :  |   |          |:     ||  |: | ::
Db    601 ------SVTEKEVSSKEEQSPVKAEVAEKQATESKPKVTKDKVVKKEIKTK--LEEKKEE 652

Qy     134 DP--DGLHQLDGTPL------TAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQ 185
           |  : : : | |||        |::  :|   | :| :|    | | |        |:
Db    653 KPKKEVVKKEDKTPLKKDEKPRKEEVKKEIKKEIKKE-ERKELKKEVKK--------ETP 703

Qy     186 AHTLEDEVAEVLQKLISKEANNYEEDPNK------------PTSWTENQAGKIPE---- 228
             : || :  :| : ||     |::| |                | | |:  :   |:
Db    704 LKDAKKEVKKEEKKEVKKE----EKEPKKEIKKISKDIKKSTPQSDTKKPSALKPKVAKK 759

Qy     229 ----KVTPMAA--IQD-GLAK-----GENDETVSNTL---------------------- 253
             |  |:||  ::| |  |      |:  |  :  :
Db    760 EESTKKEPLAAGKLKDKGKVKVIKKEGKTTEAAATAVGTAATTAAVVAAAGIAASGPVKE 819

Qy     254 -----TLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KET 298
                :| :  |  || : |   ||:    :    |: |: |:: ||          |||
Db    820 LEAERSLMSSPEDLTKDFEELKAEEIDVAKDIKPQLELIEDEEKLKETQPGEAYVIQKET 879

Qy     299 LITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL- 354
             ::              |    ||:||::  |  |    ||  :|   ||   |:| |
Db    880 EVS-------------KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFE 924

Qy     355 -----FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYL 409
                |   ||    |  :  ||| : |:: |    |         :|    :  | : : :|
Db    925 DEGAGFEESSETGDYEEKAETEEAEEPEED-GEDNASGSASKHSPTEDDESAKAEADVHL 983

Qy     410 EAIRKNI-----------EWLKK-----HDKKGNKEDY----------DLSKMRDFIN 441
           :  |:::            : |:|      :::| :||             | ::  |::
Db    984 KEKRESVVSGDDRAEEDMDDVLEKGEAEQSEEEGEEEDKAEDAREEGYEPDKTEAEDYVM 1043

Qy     442 KQADAYVEKGILDKE 456
             ||   | |: :::
Db    1044 AVADKAAEAGVTEEQ 1058
```

RESULT 4
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Dictyostelium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.

```
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discc
A;Reference number: A26655; MUID:87092266; PMID:3540939
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAR>
A;Cross-references: GB:M14628; GB:M11938; NID:g167834; PIDN:AAA33227.1; PID:g167835
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788; PMID:3901008
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116 <DEL>
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A;Reference number: S00250; MUID:88112226; PMID:2828113
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1734-1893 <WAG>
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos
F;1-818/Domain: globular head <HED>
F;89-747/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

  Query Match             7.8%;  Score 186.5;  DB 2;  Length 2116;
  Best Local Similarity   21.6%;  Pred. No. 0.29;
  Matches  108;  Conservative   86;  Mismatches  205;  Indels  101;  Gaps   17;

Qy         26 SQDKS--LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE 83
              ::|||  |  ::    |     | | | |    |      || |  |: : |      :|
Db       1023 TKDKSELLRQKKKLEEELKQVQEALAAETAAKLAQEAANKKLQGEYTEL-NEKFNSEVTA 1081

Qy         84 KEKIEKERQSIRSS--PLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQL 141
              :  :|| :::: |    ::|:|: |     |||   |  :  |  ||   ::  | |
Db       1082 RSNVEKSKKTLESQLVAVNNELDEE----KKNR---DALEKKKALDAMLEEMKDQLEST 1134

Qy        142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLI 201
              |   :  |:      :: :|:|       :|:| :     |   ||| ||| :  :|  |
Db       1135 GGEKKSLYDL------KVKQESDMEALRNQISELQSTIAKLEKIKSTLEGEVARLQGELE 1188

Qy        202 SKE--ANNYEEDPNKPTSWTENQAGKIPEKVT--------------PMAAIQDGLAKGEN 245
              :::   :| |:    |     |::: :: |:                :: :|  |::  |
Db       1189 AEQLAKSNVEKQKKKVELDLEDKSAQLAEETAAKQALDKLKKKLEQELSEVQTQLSEANN 1248

Qy        246 D-----------ETVSNTLTLTNGLERRTKTYSEDN----FEELQYFPNFYALLKSIDSE 290
                          ||  | | |    |:: |   |         ||::      : : :: |
Db       1249 KNVNSDSTNKHLETSFNNLKLELEAEQKAKQALEKKRLGLESELKH------VNEQLEEE 1302

Qy        291 KEAKEK--------ETLITIMKTLIDFVKMMVKYGTISPEEGVSYLE------------ 329
              |: ||          |  :: :|  |:       |  | :  :  | |:
Db       1303 KKQKESNEKRKVDLEKEVSELKDQIEEEVASKKAVTEAKNKKESELDEIKRQYADVVSSR 1362

Qy        330 --NLDEMIALQTKN-KLEKNATDNISKLFPAPSEKSHEETD------STKEEAAKMEKEY 380
                :::::  || || :|   | :   :|  |    |  | |      : :|| ||  |
Db       1363 DKSVEQLKTLQAKNEELRNTAEEAEGQLDRAERSKKKAEFDLEEAVKNLEEETAKKVKAE 1422

Qy        381 GSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFI 440
              ::| :  |  |       |     :| |::  | | |                ||::|  :
Db       1423 KAMKKAETDYRSTKSELDDAKNVSSEQYVQIKRLNEE--------------LSELRSVL 1467

Qy        441 NKQADAYVEKGILDKEEAEA 460
              ::||       |  |: ||:
Db       1468 -EEADERCNSAIKAKKTAES 1486


RESULT 5
B71609
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71609
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-665 <GAR>
A;Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AAC71925.1; PID:g384524
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0680w

  Query Match             7.7%;  Score 184;  DB 2;  Length 665;
  Best Local Similarity   22.0%;  Pred. No. 0.089;
  Matches  104;  Conservative   76;  Mismatches  182;  Indels  110;  Gaps   17;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEK 84
              | || |  | |   :: : | : | || | ::    ::|  ::  |           :
Db       160 GKQDISNSNAE--NKKDVKEGVKELEEKKKEEKISDDHKVEENKKS-----------DD 205

Qy         85 EKIEKERQSIRSSPLDNKLN----VEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQ 140
                |:|: ::|       :|| :    :|:|   :  :  :: |  :   ::| :|:    :
Db       206 HKVEENKKSDDHKVEENKKSDDHKIEEVKKVEEHEEDEEEDKKEKKSENKNKDENKDEND 265

Qy        141 LDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKL 200
              |   :: || |          |||    ||            |:    ||:|  |:::|
Db       266 EDNDEISDEDEVDDDVEEDKNENDDIDDDK-----------KETDKTHLEEEENEIIEKE 314

Qy        201 IS-KEANNYEEDPNKPTSW-TENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNG 258
              | |: |   :|  |  |  || :  |  ||            | :: |   :
Db       315 FSDKKKNGKNKDTKKEKSKDTEKEKSKDIEK-----------EKSKDKEKEKS------- 356

Qy        259 LERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGT 318
              : : |   :|  :|            || | ||| |||:  |   |:  |  |   |
Db       357 -KDKEKEKGKDKEKE-----------KSKDIEKE-KEKDKDIEKEKSK-DTAKEKEKDKD 402

Qy        319 ISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSH--------------E 364
              |  |:          :|  |: |   ||   ||  |      ::|                |
Db       403 IEKEKS-------KDMEKLKNKQNDEKKKDDNEKK----KNDKQDIHDDNDDENDMEEIE 451

Qy        365 ETDSTKEEAAKME-----------KEYGSLKDSTK-DDNSNPGGKTDEPKGKTEAYLEAI 412
              | |  ::|   ||           | |:   |    ::| | |  |  :| | ::|   |
Db       452 ENDDEEDEDEDMENKKKKKKGKNGNENGNENGSENGNENGNENGNENENKNESENENENE 511

Qy        413 RKNIEWLKKHDKKGNKEDYDLSKMRDF----------INKQADAYVEKGILD 454
              :|    :  ::|  |::|  :: :: :           ||| ::  :  |  :|
Db       512 NENENGNENENEKENEKDKNIKEIENVTNANKENYEKINKNSEITITKSNID 563


RESULT 6
S67593
transport protein USO1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D2552; protein YDL058w
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C;Accession: S67593; A38455; S30782
R;Bloecker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67587
A;Accession: S67593
A;Molecule type: DNA
A;Residues: 1-1790 <BLO>
A;Cross-references: EMBL:Z74106; NID:g1431058; PID:e253003; PID:g1431059; MIPS:YDL058w
A;Experimental source: strain S288C
R;Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
J. Cell Biol. 113, 245-260, 1991
A;Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transpc
A;Reference number: A38455; MUID:91185402; PMID:2010462
```

```
A;Accession: A38455
A;Molecule type: DNA
A;Residues: 1-389,'TA',392-724,'S',726-1790 <NAK>
A;Cross-references: GB:X54378; NID:g4777; PIDN:CAA38253.1; PID:g4778
A;Note: the authors translated the codon ACT for residue 768 as Ile
R;Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A;Description: An integrin analogue in Saccharomyces cerevisiae.
A;Reference number: S30782
A;Accession: S30782
A;Molecule type: DNA
A;Residues: 71-846,'E',848-923,'K',925-1252,'I',1254-1318,'V',1320-1460,'S',1462-1580,'S
A;Cross-references: EMBL:L03188
C;Genetics:
A;Gene: SGD:USO1; INT1
A;Cross-references: SGD:S0002216; MIPS:YDL058w
A;Map position: 4L
C;Keywords: coiled coil; transmembrane protein
F;326-342/Domain: transmembrane #status predicted <TM1>
F;394-410/Domain: transmembrane #status predicted <TM2>
F;617-633/Domain: transmembrane #status predicted <TM3>

  Query Match             7.6%;  Score 183.5;  DB 2;  Length 1790;
  Best Local Similarity   20.3%;  Pred. No. 0.33;
  Matches  120;  Conservative   91;  Mismatches  200;  Indels  179;  Gaps   23;

Qy         29 KSLHN--RELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI----T 82
              |||  |  ::: ||   ||  : :| |:                 | |  :    ||   |
Db        965 KSLANNYKDMQAE---NESLIKAVEE-------------SKNESSIQLSNLQNKIDSMSQ 1008

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI------DDYDSTKSGLDHKFQD--- 133
              |||  : || ||  :    |  : |:: ||   :       |:|:|  | |  | :
Db       1009 EKENFQIERGSIEKNIEQLKKTISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATT 1068

Qy        134 -DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH----- 187
               : : ::::     | |::  ::||  |:     :  |: :    |  : |:: |
Db       1069 ANDENVNKISELTKTREELEAELAA--YKNLKNELETKLETSEKALKEVKENEEHLKEEK 1126

Qy        188 -TLEDEVAEVLQKLISKEAN---------------NYEED-PNKPTSWTENQAGKIPEK 229
               || |  |  |:| |  ||                |||   ||   : | :  :: ::
Db       1127 IQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKERQYNE-EISQLNDE 1185

Qy        230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRT----------------KTYSEDN-- 270
               :|      : : | :|||          : :  | ::             |  :| |
Db       1186 ITSTQQENESIKK-KNDELEGEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEA 1244

Qy        271 ---------------FEELQYFPNF--------------------YALLKSIDSEKEAK 294
                           :|||   ||                     |  |:  :|||  :
Db       1245 SLLESIKSVESETVKIKELQDECNFKEKEVSELEDKLKASEDKNSKYLELQK-ESEKIKE 1303

Qy        295 EKETLITIMKTLIDFVKMMVK---------------------------------------- 315
               | :   | :|  :: :  : |
Db       1304 ELDAKTTELKIQLEKITNLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQ 1363

Qy        316 ------------YGTISPE--EGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEK 361
                         ||: |  | :: ||  ||:| || :|:|:    ||          |
Db       1364 AFEKERKLLNEGSSTITQEYSEKINTLE--DELIRLQNENELKAKEIDNTRSELEKVSLS 1421

Qy        362 SHEETDSTKEEAAKMEKEYGSLKDS-TKDDNSNPGGKTDEP------KGKTEAYLEAIRK 414
               : |  :  :     :: |  | ||  |::|      : |        | :  |  |:  |
Db       1422 NDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAK 1481

Qy        415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILD-KEEAEAIKR 463
               | ||| ::: :||  :| | :: : |: :: :|    :  |   | |::
Db       1482 VEEGLKKLEEESSKEKAELEKSKEMM-KKLESTIESNETELKSSMETIRK 1530


RESULT 7
A56577
microtubule-associated protein MAP 1B - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C;Accession: A56577
R;Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
Eur. J. Cell Biol. 57, 66-74, 1992
A;Title: Identification of two distinct microtubule binding domains on recombinant rat MA
A;Reference number: A56577; MUID:92347374; PMID:1639092
A;Accession: A56577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2364 <ZAU>
A;Cross-references: GB:X60550
A;Experimental source: brain
A;Note: nucleotide sequence not given; conceptual translation not complete
C;Superfamily: microtubule-associated protein MAP1B

  Query Match             7.5%;  Score 180.5;  DB 2;  Length 2364;
  Best Local Similarity   22.7%;  Pred. No. 0.65;
  Matches  120;  Conservative   77;  Mismatches  200;  Indels  131;  Gaps   22;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIK-KTYPPENKPGQSNYSFVDNLNLLKA 80
              ||  |:      ::| : |      |: :  : : | |    ::|||:        :    :
Db        454 KPLSSKSVRKESKEEAPEATKASQVEKTPKVESKEKVIVKKDKPGK--------VESKPS 505

Qy         81 ITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDP--DGL 138
              :||||    || ||   : :  |   |          |::      ||   : | :: |  :
Db        506 VTEKEVPSKEEQSPVKAEVAEKAATESKPKVTKDKVVKKEIKTKP--EEKKEEKPKKEVA 563

Qy        139 HQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHT-LEDEVAEVL 197
               : | |||  ::   |       |  :    | : |      |  | :  | |:|    ||
Db        564 KKEDKTPLKKDEKPKK-------EEAKKEIKKEIKKEEKKELKKEVKKETPLKDAKKEV- 615

Qy        198 QKLISKEANNYEEDPNK-------------PTSWTENQAG--------KIPEKVTPMAA- 235
               :|   ||    |::| |              | | |:  |          : | |  |:||
Db        616 KKDEKKEVKKEEKEPKKEIKKISKDIKKSTPLSDTKKPAALKPKVAKKEEPTKKEPIAAG 675

Qy        236 -IQD-GLAK-----GENDETVSNTL------------------------TLTNGLERRTK 264
               ::| |  |      |:  |  :  :                        :| :  |  ||
Db        676 KLKDKGKVKVIKKEGKTTEAAATAVGTAAVAAAAGVAASGPAKELEAERSLMSSPEDLTK 735

Qy        265 TYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KETLITIMKTLIDFVKMMV 314
               : |   ||:    :    |: |: |:: ||          ||| ::
Db        736 DFEELKAEEIDVAKDIKPQLELIEDEEKLKETEPGEAYVIQKETEVS------------- 782

Qy        315 KYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL------FPAPSEKSHEE 365
               |    ||:||::  |   |    ||  :|   ||   |:| |          |   ||    |
Db        783 KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFEDEGAGFEESSEAGDYE 840

Qy        366 TDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEA--YLEAIRKNI------- 416
               :  ||| : |::    :|:      |       ||   | ||  :::  |:::
Db        841 EKAETEEAEEPEEDG---EDNVSGSASKHSPTEDEEIAKAEADVHIKEKRESVASGDDRA 897

Qy        417 -----EWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE 459
                  | |:| : : ::|: :  :     :|  ||  |    || |||
Db        898 EEDMDEALEKGEAEQSEEEGEEEE-----DKAEDAREEDHEPDKTEAE 940


RESULT 8
A45605
mature-parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: A45605; A54517
R;Coppel, R.L.
Mol. Biochem. Parasitol. 50, 335-347, 1992
A;Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human ery
A;Reference number: A45605; MUID:92158014; PMID:1741020
A;Accession: A45605
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <COP>
A;Experimental source: Papua New Guinean isolate FC27
A;Note: sequence extracted from NCBI backbone (NCBIN:83648, NCBIP:83656)
R;Coppel, R.L.; Culvenor, J.G.; Bianco, A.E.; Crewther, P.E.; Stahl, H.D.; Brown, G.V.; A
Mol. Biochem. Parasitol. 20, 265-277, 1986
A;Title: Variable antigen associated with the surface of erythrocytes infected with matu
```

A;Reference number: A54517; MUID:87014571; PMID:3531849
A;Accession: A54517
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 222-443,'K',445 <CO2>
A;Cross-references: GB:M15319; NID:g160060; PID:g552170
C;Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology
C;Keywords: surface antigen; tandem repeat

Query Match 7.5%; Score 179.5; DB 2; Length 1526;
Best Local Similarity 18.7%; Pred. No. 0.42;
Matches 107; Conservative 101; Mismatches 182; Indels 181; Gaps 24;

```
Qy         25 GSQDKSLHNRELS--AERPLNEQIAEAEEDKIKKTYPPENKPGQSN---------- 68
Db        400 GASENSEDPKKLTEQEENGTKESSEETKDDK-----PEENEKKADNKKKSKKKKKSFFQM 454

Qy         69 --YSFVDNLNL-----LKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYD 121
Db        455 LGCNFLCNKNIETDDEEETLVVKDDAKKKHKFLREA------NTEKNDNEKKDKLLGEGD 508

Qy        122 ----------------STKSGLDHKFQDDPDGL----------------HQLDGTPL 146
Db        509 KEDVKEKNDEQKDKVLGEGDKEDVKEKNDEQKDKVLGEGDKEDVKEKNDGKKDKVIGSEK 568

Qy        147 TAEDIVHKIAARI-----------YEENDRAVFDKIVSKLLNLGLITESQAHTLEDEV-- 193
Db        569 TQKEIKEKVEKRVKKKCKKKVKKGIKENDTEGNDKVKGPEIIIEEVKEEIKKQVEDGIKE 628

Qy        194 -----------AEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTP------MAAI 236
Db        629 NDTEGNDKVKGPEIITEEVKEEIKKQVEDGIKENDTEGNDKVKGPEIITEEVKEEIKKQV 688

Qy        237 QDGLAKG--ENDETVSNTLTLTNGLERRTK-----------TYSEDNF-------EELQY 276
Db        689 EEGIKENDTEGNDKVKGPEIITEEVKEEIKKQVEEGVKENDTESKDKVIGQEIITEEVK- 747

Qy        277 FPNFYALLKSIDSEKEAKEKETLITI------MKTLIDFVKMMVKYGTISPEEGV--SYL 328
Db        748 --------KEIEKQEEKGNKENILEIKDIVIGQEVIIEEVKKVIKK---KVEKGIKENHT 796

Qy        329 ENLD-----EMIALQTKNKLEKNATDNISKLFPAPSEK----------SHEETDSTKEEA 373
Db        797 ESKDKVIGQEIIVEEVKEEIEKQVEEGIKENDTESKDKVIGEEVIKGDVNEEGPENKDKV 856

Qy        374 AKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL 433
Db        857 TKQEK----VKEVKK-----------EVKKK-------VKKRV---KKRNNKNERKDNVI 891

Qy        434 SK--MRDFINKQADAYVEKGILDKEEAEAIK 462
Db        892 GKEIMKEDVNEKDTANKDKEIEQEKEKEEVK 922
```

RESULT 9
T51505
hypothetical protein F5E19_70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51505
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-853 <SAT>
A;Cross-references: EMBL:AL391147
A;Experimental source: cultivar Columbia; BAC clone F5E19
C;Genetics:
A;Map position: 5
A;Introns: 6/2; 79/3
A;Note: F5E19_70

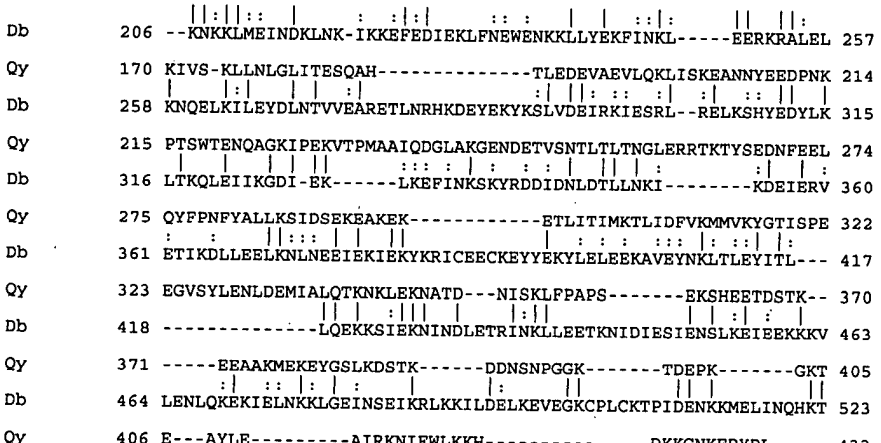
Query Match 7.5%; Score 179; DB 2; Length 853;
Best Local Similarity 20.9%; Pred. No. 0.21;
Matches 102; Conservative 81; Mismatches 206; Indels 100; Gaps 19;

```
Qy         16 PIQAFPKPGGSQDKSLHNRELSAER-----PLNEQIAEAEEDKIKKTYPPEN-------- 62
Db         45 PSTTTPHSRLSLDRSSPNSKSSVERRSPKLPTPPEKSQARVAAVKGTESPQTTTRLSQIK 104

Qy         63 ---KPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDD 119
Db        105 EDLKKANERISSLEK-DKAKALDELKQAKKEAEQV-TLKLDDALKAQ--KHVEENSEIEK 160

Qy        120 YDSTKSGLDHKFQDDPDGL----------HQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
Db        161 FQAVEAGIE-AVQNNEEELKKELETVKNQHASDSAALVA---VRQELEKINEEL-AAAFD 215

Qy        170 KIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEK 229
Db        216 -AKSKAL-------SQA----------EDASKTAEIHAEKVDILSSE 244

Qy        230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDS 289
Db        245 LTRLKALLD----------STREKTAISDNEMVAKLEDEIVVLKRDLES 283

Qy        290 ----EKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEK 345
Db        284 ARGFEAEVKEKEMIVEKLNVDLEAAKMAESNAHSLSNEWQSKAKELEEQ--LEEANKLER 341

Qy        346 NAT---DNISKLFPAPSEKSHE-ETDST--KEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
Db        342 SASVSLESVMKQLEGSNDKLHDTETEITDLKERIVTLETTVAKQKEDLEVSEQRL-GSVE 400

Qy        400 EPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE 459
Db        401 EEVSKNEKEVEKLKSELETVKEEKNRALKKEQDATSRVQRLSEEKSKLLSDLESSKEEEE 460

Qy        460 AIKRIYSSL 468
Db        461 KSKKAMESL 469
```

RESULT 10
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: A64465
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: A64465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1005 <BUL>
A;Cross-references: GB:U67572; GB:L77117; NID:g1591958; PIDN:AAB99331.1; PID:g1591962; T
C;Genetics:
A;Map position: REV1273394-1270377
C;Superfamily: hypothetical protein MJ1322

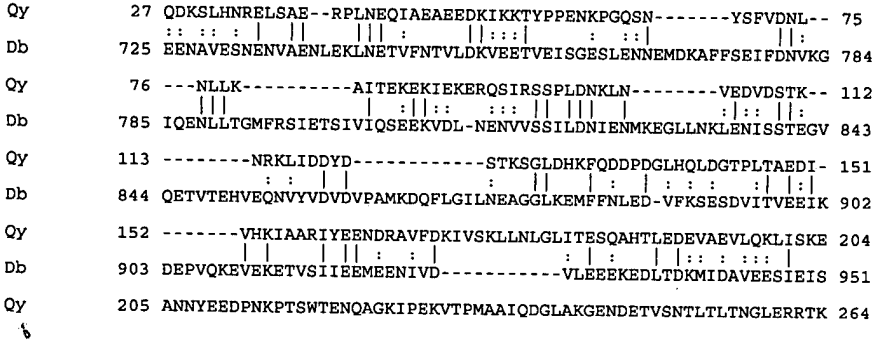
Query Match 7.5%; Score 179; DB 2; Length 1005;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 116; Conservative 80; Mismatches 164; Indels 160; Gaps 27;

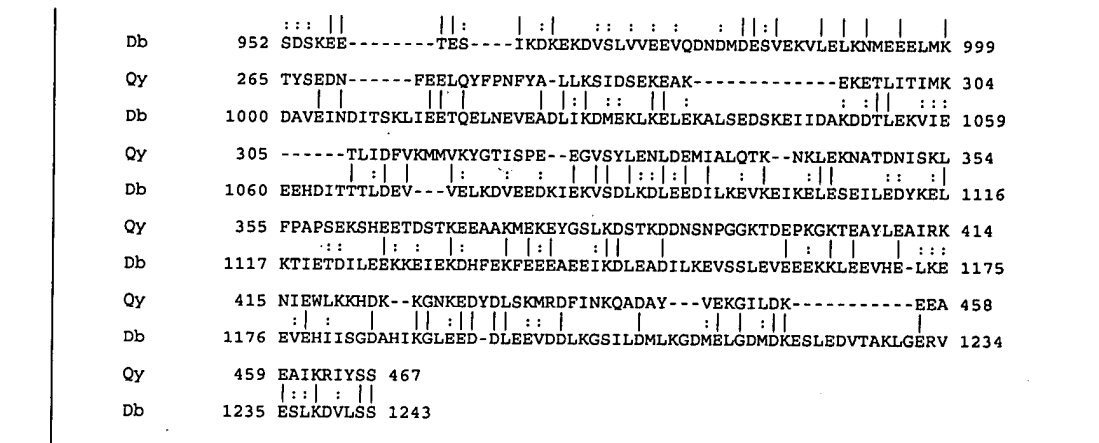
```
Qy         52 DKIKKTYPPENKPGQSNYSFVDNLNLLKA-ITEKEKIEKERQSIRSSPLDNKLN-VEDVD 109
Db        159 DEFEKCY---QKMGEIVKEYEKRLERIEGELNYKENYEKE--------LKNKMSQLEE-- 205

Qy        110 STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
```



RESULT 11
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71603
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71603
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1558 <GAR>
A;Cross-references: GB:AE001424; GB:AE001362; NID:g3845307; PIDN:AAC71972.1; PID:g384530
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0915w

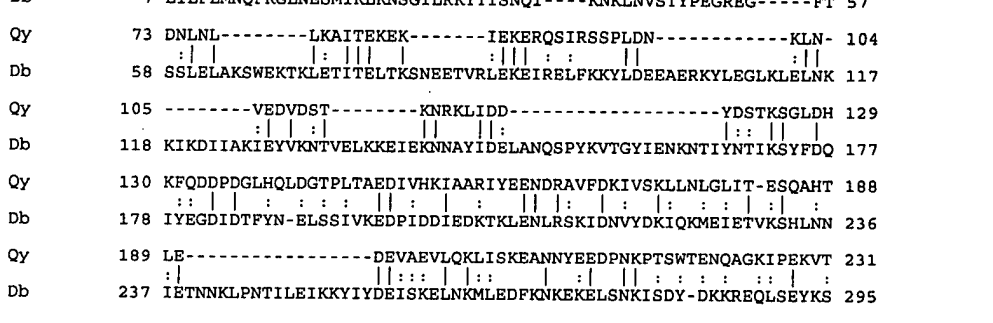
Query Match 7.5%; Score 179; DB 2; Length 1558;
Best Local Similarity 21.5%; Pred. No. 0.45;
Matches 118; Conservative 96; Mismatches 197; Indels 138; Gaps 25;





RESULT 12
T28677
rhoptry protein - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28677; C45521
R;Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A;Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A;Reference number: Z20508; MUID:95021522; PMID:7935623
A;Accession: T28677
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2269 <KEE>
A;Cross-references: EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304.1
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cop
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: C45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2131-2269 <KE2>
A;Cross-references: GB:M34283

Query Match 7.5%; Score 179; DB 2; Length 2269;
Best Local Similarity 21.1%; Pred. No. 0.73;
Matches 113; Conservative 85; Mismatches 193; Indels 144; Gaps 23;



```
Qy        232 PMAAIQDGLAKGENDET-VSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSE 290
              |   |::       | :| | ||       |   : | : | | :    |      :  | ||
Db        296 KMLEIRNHY----NSQTNVDNTKE-----EEAKQNYDKSN-EHMTTIPTNEDEISKIISE 345

Qy        291 KEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDN 350
              :   |: ::: : | ||| |   ||       |   : |  |:: |  :  :|:|
Db        346 VKTM-KDEILSKVNTYIDFNK---KYKETVNSEHSQFTELTDKIKAEVSDKELKK----- 396

Qy        351 ISKLFPAPSEKSHEETDS-TKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYL 409
                        |:|  :  |   |    :|||          | |   | ||       |:
Db        397 --------CEQSFNDNKSLINETKNSIEKEY---------QNINTLKKVDE-------YI 432

Qy        410 EAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRI 464
              : :   | : |   |      : ::| :|:       |   :||  |  ::|
Db        433 KVCKSTKESITKFSSK-------QTILKDMLNQNIKTVKETNSIDKSYIEKFEQI 480


RESULT 13
A43336
microtubule-vesicle linker CLIP-170 - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C;Accession: A43336
R;Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A;Title: CLIP-170 links endocytic vesicles to microtubules.
A;Reference number: A43336; MUID:92405160; PMID:1356075
A;Accession: A43336
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1392 <PIE>
A;Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

  Query Match             7.4%;  Score 178.5;  DB 2;  Length 1392;
  Best Local Similarity   22.4%;  Pred. No. 0.42;
  Matches  111;  Conservative   79;  Mismatches  182;  Indels  123;  Gaps   20;

Qy         34 RELSAE-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL--------NLLKA 80
              |: |:|     : | :|: || | :||     :|       |    |         || :
Db        734 RKASSEGKSEMKKLRQQL-EAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTNLQEN 792

Qy         81 ITE----KEKIEKERQ-------------SIRSSPLD--NKL-------NVEDVDSTKN 113
              ::|    || :||| |              |:: |  :  |||        |:   |  |
Db        793 LSEVSQVKETLEKELQILKEKFAEASEEAVSVQRSMQETVNKLHQKEEQFNMLSSDLEKL 852

Qy        114 RKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS 173
              |: : |       :: ||::  :   ||             | |:   |||       ::
Db        853 RENLAD-------MEAKFREKDEREEQL-------------IKAKEKLEND-------IA 885

Qy        174 KLLNLGLITESQAHTLEDEV------AEVLQKLISKEANNYE------EDPNKPTSWTEN 221
              ::: :     ||   : ||:       | ||  ::|    |          ||       ::
Db        886 EIMKMSGDNSSQLTKMNDELRLKERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQ 945

Qy        222 QAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFY 281
              :| |  |:            | | :  :|:       ||:: :| | :  :||:
Db        946 EAAKKHEE-----------EKKELERKLSD-------LEKKMET-SHNQCQELK------ 980

Qy        282 ALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKN 341
              |  :   || : | :| |  : |||:|    :      : ||    |: |:|:     |
Db        981 ARYERATSETKTKHEEILQNLQKTLLDTEDKLKG----AREENSGLLQELEELRKQAEKA 1036

Qy        342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
              |  : | | :  :      ||:      : ||:  ||:: |  :||::    |      |:
Db       1037 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNAKLQNELDTLKENNL-KNVEELNKSK 1095

Qy        400 EPKGKTEAYLEAIRKNIEWLKKH----------DKKGNKEDYDLSKMRDFINKQADAYV 448
              |        :|  || || ||:            ::  |   :| : || :
Db       1096 ELLTVENQKMEEFRKEIETLKQAAAQKSQQLSALQEENVKLAEELGRSRDEVTSHQKLEE 1155

Qy        449 EKGILDKEEAEAIKR 463
              |: :|: :  |  ||
Db       1156 ERSVLNNQLLEMKKR 1170


RESULT 14
S22695
restin - human
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999
C;Accession: S22695; S19853
R;Bilbe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.; S
EMBO J. 11, 2103-2113, 1992
A;Title: Restin: a novel intermediate filament-associated protein highly expressed in the
A;Reference number: S22695; MUID:92289675; PMID:1600942
A;Accession: S22695
A;Molecule type: mRNA
A;Residues: 1-1427 <BIL>
A;Cross-references: EMBL:X64838; NID:g35998; PIDN:CAA46050.1; PID:g35999
C;Keywords: cytoskeleton

  Query Match             7.4%;  Score 178.5;  DB 2;  Length 1427;
  Best Local Similarity   22.4%;  Pred. No. 0.43;
  Matches  111;  Conservative   79;  Mismatches  182;  Indels  123;  Gaps   20;

Qy         34 RELSAE-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL--------NLLKA 80
              |: |:|     : | :|: || | :||     :|       |    |         || :
Db        769 RKASSEGKSEMKKLRQQL-EAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTNLQEN 827

Qy         81 ITE----KEKIEKERQ-------------SIRSSPLD--NKL-------NVEDVDSTKN 113
              ::|    || :||| |              |:: |  :  |||        |:   |  |
Db        828 LSEVSQVKETLEKELQILKEKFAEASEEAVSVQRSMQETVNKLHQKEEQFNMLSSDLEKL 887

Qy        114 RKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS 173
              |: : |       :: ||::  :   ||             | |:   |||       ::
Db        888 RENLAD-------MEAKFREKDEREEQL-------------IKAKEKLEND-------IA 920

Qy        174 KLLNLGLITESQAHTLEDEV------AEVLQKLISKEANNYE------EDPNKPTSWTEN 221
              ::: :     ||   : ||:       | ||  ::|    |          ||       ::
Db        921 EIMKMSGDNSSQLTKMNDELRLKERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQ 980

Qy        222 QAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFY 281
              :| |  |:            | | :  :|:       ||:: :| | :  :||:
Db        981 EAAKKHEE-----------EKKELERKLSD-------LEKKMET-SHNQCQELK------ 1015

Qy        282 ALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKN 341
              |  :   || : | :| |  : |||:|    :      : ||    |: |:|:     |
Db       1016 ARYERATSETKTKHEEILQNLQKTLLDTEDKLKG----AREENSGLLQELEELRKQADKA 1071

Qy        342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
              |  : | | :  :      ||:      : ||:  ||:: |  :||::    |      |:
Db       1072 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNAKLQNELDTLKENNL-KNVEELNKSK 1130

Qy        400 EPKGKTEAYLEAIRKNIEWLKKH----------DKKGNKEDYDLSKMRDFINKQADAYV 448
              |        :|  || || ||:            ::  |   :| : || :
Db       1131 ELLTVENQKMEEFRKEIETLKQAAAQKSQQLSALQEENVKLAEELGRSRDEVTSHQKLEE 1190

Qy        449 EKGILDKEEAEAIKR 463
              |: :|: :  |  ||
Db       1191 ERSVLNNQLLEMKKR 1205


RESULT 15
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96796
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
```

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1871 <STO>
A;Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141
C;Genetics:
A;Gene: F28O16.15
A;Map position: 1

Query Match 7.4%; Score 177; DB 2; Length 1871;
Best Local Similarity 20.7%; Pred. No. 0.71;
Matches 116; Conservative 85; Mismatches 212; Indels 148; Gaps 22;

```
Qy        12 VLVLPIQAFPKP--GGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNY 69
             |||   : :||    | :| : |  |    |  : :     |||  ||    | :
Db      1052 VLVEEEETYPKDKHTGGEDHNDHKEEEQKENVIAKAELNTEEDSFKKVEEIEKQ------ 1105

Qy        70 SFVDNLNLLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKN---RKLIDDYDSTKSG 126
               |:  | ::: : :: | | :  ::  ::    ||    ||:    || :  |    |
Db      1106 ---DHGELKRSMVQAKRQETEEKD-KTRAMEKNETVERRKQTKDGSLGKLREGEDPELGG 1161

Qy       127 LDHKFQDDPDGLHQLDGTPLT--AEDIVHKIAARIYEENDRAVFD----KIVSKLLNLGL 180
               |: : : | : :|  | ::   | :  |    |    |    |    :  ||   :
Db      1162 --HERRGEEDRIEELVETEISDHKEKVKKKDEDYILRSQDTGKVDLGERERRSKQRKIHK 1219

Qy       181 ITESQAHTLEDEVAEVLQKLISKEAN-------NYEEDPNKPTSWTENQAGKIPEKVTP- 232
             | :    ||| ||    ::|:  |          ||:  |       :  ||||   |
Db      1220 SVEDEIGDQEDEDAEEAAAVVSRNENGSSRKVQTIEEESEK-----HKEQNKIPETSNPE 1274

Qy       233 ----------------MAAIQDGLAKGEN----------DETVSNTLTLTNGLERRTKT 265
                           | :|:   | ||          :|     :|  | | :| ||
Db      1275 VNEEDEERVVEKETKEVEAHVQELEGKTENCKDDDGEGRREERGKQGMTAENMLRQRFKT 1334

Qy       266 YSEDNFEELQYFPNFYALLKSID--SEKEAKEKETLITIMKTLIDFVKMMVKYGTI---- 319
             |:|              ::: |    |:|  ||::      ::    ||:: : |::
Db      1335 KSDD-----------GIVRKIQETKEEEPDEKKS----QESSSHVVKLVAEDGSLRNGL 1378

Qy       320 ---SPEEGVSYLENLDEMIALQTKNKLEKNATD--NISKLFPAPSEKSHEETDST----- 369
                |  || :  |||    :   |: |   :  |   :    ::|: ||
Db      1379 EFSEKESTVSKMLKLDESKEKEEHKKIRKPTEERSNAPVIEKQGNKKNAEEEMQDKIDRR 1438

Qy       370 -KEEAAKMEKEYGSLKD--------------------------STKDDNSNPGGKTD-- 399
             | :  | :: || |::                          |||:   |   ::
Db      1439 GKNQEIKGQEPYGVLRNGEHDKITEYHRGEEKGTAENVSSTKIQQTKDELEKPRKPSEIS 1498

Qy       400 ----------------EPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLS-KMRDFIN 441
                           | ||  :|   | :  |: :   :  |  ||:| :| ::|: :
Db      1499 ENHNIHEFMDSSQSQDIEEKGSDQAEKYAKQNKIQEVMNDEDK--KEEYHISERVRNEMA 1556

Qy       442 K---QADAYVEKGILDKEEAE 459
             |   | ::    |   | | |
Db      1557 KRILQVESKANDGSSKKNETE 1577
```

Search completed: January 2, 2004, 11:02:28
Job time : 24 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 17 Seconds
(without alignments)
1294.618 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2395	99.8	468	1	SG3_HUMAN	Q8wxd2 homo sapien
2	2102.5	87.6	471	1	SG3_MOUSE	P47867 mus musculu
3	2096.5	87.4	471	1	SG3_RAT	P47868 rattus norv
4	191	8.0	2464	1	MAPB_MOUSE	P14873 mus musculu
5	186.5	7.8	2116	1	MYS2_DICDI	P08799 dictyosteli
6	183.5	7.6	1790	1	USO1_YEAST	P25386 saccharomyc
7	180.5	7.5	2459	1	MAPB_RAT	P15205 rattus norv
8	179	7.5	1005	1	RA50_METJA	Q58718 methanococc
9	178.5	7.4	1427	1	REST_HUMAN	P30622 homo sapien
10	178	7.4	882	1	RA50_PYRFU	P58301 pyrococcus
11	172.5	7.2	1164	1	BAG_STRAG	P27951 streptococc
12	169	7.0	700	1	TRDN_CANFA	P82179 canis famil
13	167.5	7.0	957	1	KF5C_HUMAN	O60282 homo sapien
14	166.5	6.9	1875	1	MLP1_YEAST	Q02455 saccharomyc
15	166	6.9	2663	1	CENE_HUMAN	Q02224 homo sapien
16	166	6.9	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
17	163.5	6.8	956	1	KF5C_MOUSE	P28738 mus musculu
18	163	6.8	1251	1	RBP2_PLAVB	Q00799 plasmodium
19	161.5	6.7	1360	1	CING_XENLA	Q9ptd7 xenopus lae
20	161.5	6.7	1969	1	MYSA_CAEEL	P12844 caenorhabdi
21	161	6.7	852	1	RA50_THEMA	Q9x1x1 thermotoga
22	159.5	6.6	997	1	SCP1_RAT	Q03410 rattus norv
23	159.5	6.6	1978	1	MYHB_CHICK	P10587 gallus gall
24	159	6.6	1208	1	PCP1_SCHPO	Q92351 schizosacch
25	158.5	6.6	1526	1	MYS2_SCHPO	Q9usi6 schizosacch
26	158	6.6	1972	1	MYHB_RABIT	P35748 oryctolagus
27	157.5	6.6	1972	1	MYHB_HUMAN	P35749 homo sapien
28	157.5	6.6	2230	1	GOG4_HUMAN	Q13439 homo sapien
29	157	6.5	727	1	MFP1_ARATH	Q9lw85 arabidopsis
30	157	6.5	944	1	NUF1_YEAST	P32380 saccharomyc
31	157	6.5	2017	1	MYSN_DROME	Q99323 drosophila
32	156.5	6.5	1087	1	AKA9_RABIT	Q28628 oryctolagus
33	156.5	6.5	2468	1	MAPB_HUMAN	P46821 homo sapien
34	156	6.5	705	1	TRDN_RABIT	Q28820 oryctolagus
35	155.5	6.5	1130	1	YL17_CAEEL	Q11102 caenorhabdi
36	155.5	6.5	5085	1	PCLO_RAT	Q9jks6 rattus norv
37	154.5	6.4	1539	1	Y373_HUMAN	O15078 homo sapien
38	154	6.4	1433	1	REST_CHICK	O42184 gallus gall
39	154	6.4	1957	1	SPOF_SCHPO	Q10411 schizosacch
40	154	6.4	1962	1	MYSA_DROME	P05661 drosophila
41	153.5	6.4	1976	1	MYHA_RAT	Q9jlt0 rattus norv
42	153.5	6.4	2104	1	MYS3_SCHPO	O14157 schizosacch
43	152.5	6.4	2022	1	ANT1_ONCVO	P21249 onchocerca
44	152.5	6.4	8545	1	ANC1_CAEEL	Q9n4m4 caenorhabdi
45	152	6.3	667	1	CYL1_BOVIN	P35662 bos taurus

ALIGNMENTS

```
RESULT 1
SG3_HUMAN
ID   SG3_HUMAN      STANDARD;      PRT;   468 AA.
AC   Q8WXD2; Q96C83; Q96GE8; Q9Y6G7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Secretogranin III precursor (SgIII).
GN   SCG3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pituitary;
RA   Song H., Peng Y., Huang Q., Dai M., Mao Y., Zhang Q., Mao M., Fu G.,
RA   Luo M., Chen J., Hu R.;
RT   "Human secretogranin III mRNA, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=22093564; PubMed=12098761;
RA   Rong Y.P., Liu F., Zeng L.C., Ma W.J., Wei D.Z., Han Z.G.;
RT   "Cloning and characterization of a novel human secretory protein:
RT   secretogranin III.";
RL   Acta Biochim. Biophys. Sin. 34:411-417(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC       secretory granules.
CC   -!- TISSUE SPECIFICITY: Expressed in brain, heart, kidney, liver and
CC       skeletal muscle.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF078851; AAD44483.1; -.
DR   EMBL; AF453583; AAL67431.1; -.
DR   EMBL; BC014539; AAH14539.1; -.
DR   EMBL; BC009511; AAH09511.1; ALT_INIT.
DR   Genew; HGNC:13707; SCG3.
KW   Signal; Cleavage on pair of basic residues; Polymorphism.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    468       SECRETOGRANIN III.
FT   VARIANT     125    125       S -> N (in dbSNP:2305710).
FT                                /FTId=VAR_013827.
FT   CONFLICT     79     79       K -> R (IN REF. 1).
FT   CONFLICT    167    167       A -> V (IN REF. 1 AND 2).
FT   CONFLICT    272    274       EEL -> RDF (IN REF. 1).
SQ   SEQUENCE   468 AA;  52977 MW;  633A87692A783808 CRC64;

  Query Match             99.8%;  Score 2395;  DB 1;  Length 468;
  Best Local Similarity   99.8%;  Pred. No. 6e-108;
  Matches  467;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGFLGTGTWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPP 60

Qy         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDY 120

Qy        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGL 180
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        121 DSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAAFDKIVSKLLNLGL 180

Qy        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240

Qy        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLI 300

Qy        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSE 360

Qy        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLK 420

Qy        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468


RESULT 2
SG3_MOUSE
ID   SG3_MOUSE      STANDARD;      PRT;   471 AA.
AC   P47867;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Secretogranin III precursor (SgIII).
GN   SCG3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95001263; PubMed=7917832;
RA   Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA   Sutcliffe J.G.;
RT   "Primary structure of mouse secretogranin III and its absence from
RT   deficient mice.";
RL   J. Mol. Neurosci. 4:225-233(1993).
CC   -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC       secretory granules.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U02982; AAA56636.1; -.
DR   MGD; MGI:103032; Scg3.
KW   Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    471       SECRETOGRANIN III.
SQ   SEQUENCE   471 AA;  53326 MW;  FC1E9C381AFCA564 CRC64;

  Query Match             87.6%;  Score 2102.5;  DB 1;  Length 471;
  Best Local Similarity   88.3%;  Pred. No. 5.8e-94;
  Matches  416;  Conservative   17;  Mismatches   35;  Indels    3;  Gaps    1;

Qy          1 MGFLGTGTWILVLVL---PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKT 57
              |||| ||:|||||||   |||||||| ||||||||||||||||||||||||| |||||
Db          1 MGFLWTGSWILVLVLNSGPIQAFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKA 60

Qy         58 YPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI 117
              :| |:|| :|||| |||||||:|||||| :||||||||| | ||:||||| |||||||||
Db         61 FPSESKPSESNYSSVDNLNLLRAITEKETVEKERQSIRSPPFDNQLNVEDADSTKNRKLI 120

Qy        118 DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLN 177
              |:|||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db        121 DEYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLN 180

Qy        178 LGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
              |||||||||||||||| |||||||||||||  :|||| |||| |||||||||:||:|
Db        181 LGLITESQAHTLEDEVAEALQKLISKEANNYEETLDKPTSRTENQDGKIPEKVTPVAAVQ 240

Qy        238 DGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              ||    ||||||||||:|||||| : ||:|||||||||||||| ||||||||||||
Db        241 DGFTNRENDETVSNTLTLSNGLERRTNPHREDDFEELQYFPNFYALLTSIDSEKEAKEKE 300

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||: ||||||
Db        301 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDETIALQTKNKLEKNTTDSKSKLFPA 360

Qy        358 PSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIE 417
              | ||| ||||||||||||||||||||||||||||||||| ||||||  ||||||||||||||
Db        361 PPEKSQEETDSTKEEAAKMEKEYGSLKDSTKDDNSNLGGKTDEATGKTEAYLEAIRKNIE 420

Qy        418 WLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||:||||||||||||||:|||||||||||||||| |||||||||
Db        421 WLKKHNKKGNKEDYDLSKMRDFINQQADAYVEKGILDKEEANAIKRIYSSL 471


RESULT 3
SG3_RAT
ID   SG3_RAT        STANDARD;      PRT;   471 AA.
AC   P47868;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Secretogranin III precursor (SgIII) (1B1075).
GN   SCG3.
```

```
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95001263; PubMed=7917832;
RA   Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA   Sutcliffe J.G.;
RT   "Primary structure of mouse secretogranin III and its absence from
RT   deficient mice.";
RL   J. Mol. Neurosci. 4:225-233(1993).
RN   [2]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=90376160; PubMed=2204688;
RA   Ottiger H.-P., Battenberg E.F., Tsou A.-P., Bloom F.E.,
RA   Sutcliffe J.G.;
RT   "1B1075: a brain- and pituitary-specific mRNA that encodes a novel
RT   chromograNIN / SECretogranin-like component of intracellular
RT   vesicles.";
RL   J. Neurosci. 10:3135-3147(1990).
CC   -!- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC       secretory granules.
CC   -!- TISSUE SPECIFICITY: BRAIN- AND PITUITARY-SPECIFIC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U02983; AAA56637.1; -.
KW   Signal; Cleavage on pair of basic residues.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    471       SECRETOGRANIN III.
SQ   SEQUENCE   471 AA;  53183 MW;  977B3F8885F33223 CRC64;

  Query Match             87.4%;  Score 2096.5;  DB 1;  Length 471;
  Best Local Similarity   87.9%;  Pred. No. 1.1e-93;
  Matches  414;  Conservative   18;  Mismatches   36;  Indels    3;  Gaps    1;

Qy          1 MGFLGTGTWILVLVL---PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKT 57
              |||| ||:|||||||   |||||||| ||||||||||||||||||||||||| ||||||
Db          1 MGFLWTGSWILVLVLNSGPIQAFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKT 60

Qy         58 YPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI 117
              || |:|| :||:| ||||||||||||| :|| :||||||| ||:|||:| ||||||||
Db         61 YPSESKPSESNFSSVDNLNLLKAITEKETVEKAKQSIRSSPFDNRLNVDDADSTKNRKLT 120

Qy        118 DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLN 177
              |:|||||||| | |||||||||||||||||||||||| |||||||| ||||||||||
Db        121 DEYDSTKSGLDRKVQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLN 180

Qy        178 LGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
              |||||||||||||||| |||||||||||||| | |||| |||| |||||||||:|| |
Db        181 LGLITESQAHTLEDEVAEALQKLISKEANNYEEAPEKPTSRTENQDGKIPEKVTPVAATQ 240

Qy        238 DGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              ||    |||:||||||||:|||||||  : :|:|||||||||||||| ||||||||||||
Db        241 DGFTNRENDDTVSNTLTLSNGLERRTNPHRDDDFEELQYFPNFYALLTSIDSEKEAKEKE 300

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              |||||||||||||||||||||||||||||||||||| ||||||||||| ||: ||||||
Db        301 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDETIALQTKNKLEKNTTDSKSKLFPA 360

Qy        358 PSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIE 417
              | |||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db        361 PPEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNLGGKTDEAKGKTEAYLEAIRKNIE 420

Qy        418 WLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
                  |||||:|||||||||||||||:|||||||||||||||| |||||||||
Db        421 WLKKHNKKGNKEDYDLSKMRDFINQQADAYVEKGILDKEEANAIKRIYSSL 471


RESULT 4
MAPB_MOUSE
ID   MAPB_MOUSE     STANDARD;      PRT;  2464 AA.
AC   P14873;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE   [Contains: MAP1 light chain LC1].
GN   MAP1B OR MTAP1B OR MTAP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DOMAIN.
RC   STRAIN=Swiss Webster; TISSUE=Brain;
RX   MEDLINE=90094539; PubMed=2480963;
RA   Noble M., Lewis S.A., Cowan N.J.;
RT   "The microtubule binding domain of microtubule-associated protein
RT   MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT   and tau.";
RL   J. Cell Biol. 109:3367-3376(1989).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC       OF MAP1B.
CC   -!- SIMILARITY: TO MAP1A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51396; CAA35761.1; -.
DR   PIR; S07549; QRMSP1.
DR   MGD; MGI:1306778; Mtap1b.
DR   GO; GO:0016358; P:dendrite morphogenesis; IMP.
DR   GO; GO:0001578; P:microtubule bundling; IMP.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 7.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?   2464       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1874   1890       MAP1B 1.
FT   REPEAT     1891   1907       MAP1B 2.
FT   REPEAT     1908   1924       MAP1B 3.
FT   REPEAT     1925   1941       MAP1B 4.
FT   REPEAT     1942   1958       MAP1B 5.
FT   REPEAT     1959   1975       MAP1B 6.
FT   REPEAT     1993   2009       MAP1B 7.
FT   REPEAT     2010   2026       MAP1B 8.
FT   REPEAT     2027   2043       MAP1B 9.
FT   REPEAT     2044   2060       MAP1B 10.
FT   DOMAIN      589    787       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
```

```
FT                                KKEE AND KKEI/V REPEATS).
SQ   SEQUENCE   2464 AA;  270408 MW;  FBD3DD99CFDBDA87 CRC64;

  Query Match             8.0%;  Score 191;  DB 1;  Length 2464;
  Best Local Similarity   21.3%;  Pred. No. 0.098;
  Matches  118;  Conservative   85;  Mismatches  186;  Indels  166;  Gaps   24;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEA------EEDKIKKTYP--PENKPGQSNYSFVD 73
              ||  |:     ::| : |     |: :       |:  :||  |   |:||
Db        550 KPVASKSVRKESKEETPEVTKTSQVEKTPKVESKEKVLVKKDKPVKTESKP--------- 600

Qy         74 NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQD 133
                 ::||||   || ||   : :  |   |       |::     ||  |: | ::
Db        601 ------SVTEKEVSSKEEQSPVKAEVAEKQATESKPKVTKDKVVKKEIKTK--LEEKKEE 652

Qy        134 DP--DGLHQLDGTPL------TAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQ 185
              |  : : : | |||        |::  :|   | :| :|    | | |       |:
Db        653 KPKKEVVKKEDKTPLKKDEKPRKEEVKKEIKKEIKKE-ERKELKKEVKK--------ETP 703

Qy        186 AHTLEDEVAEVLQKLISKEANNYEEDPNK-------------PTSWTENQAGKIPE---- 228
                : || :  :| : ||    |::| |             | | |:  :   |:
Db        704 LKDAKKEVKKEEKKEVKKE----EKEPKKEIKKISKDIKKSTPQSDTKKPSALKPKVAKK 759

Qy        229 ----KVTPMAA--IQD-GLAK-----GENDETVSNTL---------------------- 253
                  |  |:||  ::| |  |     |:  |  :  :
Db        760 EESTKKEPLAAGKLKDKGKVKVIKKEGKTTEAAATAVGTAATTAAVVAAAGIAASGPVKE 819

Qy        254 -----TLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KET 298
                  :| :  |  || : |   ||:    :    |: |: |:: ||          |||
Db        820 LEAERSLMSSPEDLTKDFEELKAEEIDVAKDIKPQLELIEDEEKLKETQPGEAYVIQKET 879

Qy        299 LITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL- 354
               ::             |    ||:||::  |   |    ||  :|   ||   |:| |
Db        880 EVS-------------KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFE 924

Qy        355 -----FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYL 409
                  |   ||    |  :  ||| : |:: |     |       :|    :  | : : :|
Db        925 DEGAGFEESSETGDYEEKAETEEAEEPEED-GEDNASGSASKHSPTEDDESAKAEADVHL 983

Qy        410 EAIRKNI------------EWLKK-----HDKKGNKEDY-----------DLSKMRDFIN 441
              :  |:::             : |:|      :::| :||           | ::  |::
Db        984 KEKRESVVSGDDRAEEDMDDVLEKGEAEQSEEEGEEEDKAEDAREEGYEPDKTEAEDYVM 1043

Qy        442 KQADAYVEKGILDKE 456
                ||   | |: :::
Db       1044 AVADKAAEAGVTEEQ 1058


RESULT 5
MYS2_DICDI
ID   MYS2_DICDI     STANDARD;      PRT;  2116 AA.
AC   P08799;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Myosin II heavy chain, non muscle.
GN   MHCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87092266; PubMed=3540939;
RA   Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT   "Conserved protein domains in a myosin heavy chain gene from
RT   Dictyostelium discoideum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN   [2]
RP   PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC   STRAIN=AX2;
RX   MEDLINE=90353583; PubMed=2387408;
RA   Lueck-Vielmeter D., Schleicher M., Grabatin B., Wippler J.,
RA   Gerisch G.;
RT   "Replacement of threonine residues by serine and alanine in a
RT   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL   FEBS Lett. 269:239-243(1990).
RN   [3]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=88112226; PubMed=2828113;
RA   Wagle G., Noegel A., Scheel J., Gerisch G.;
RT   "Phosphorylation of threonine residues on cloned fragments of the
RT   Dictyostelium myosin heavy chain.";
RL   FEBS Lett. 227:71-75(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345066; PubMed=7619795;
RA   Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA   Rayment I.;
RT   "X-ray structures of the myosin motor domain of Dictyostelium
RT   discoideum complexed with MgADP.BeFx and MgADP.AlF4-.";
RL   Biochemistry 34:8960-8972(1995).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345067; PubMed=7619796;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT   truncated head of Dictyostelium discoideum myosin to 2.7-A
RT   resolution.";
RL   Biochemistry 34:8973-8981(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=96206189; PubMed=8611530;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL   Biochemistry 35:5404-5417(1996).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX   MEDLINE=97452580; PubMed=9305951;
RA   Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT   "X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes
RT   of the Dictyostelium discoideum myosin motor domain.";
RL   Biochemistry 36:11619-11628(1997).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=98070605; PubMed=9405148;
RA   Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT   "X-ray crystal structure and solution fluorescence characterization
RT   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT   Dictyostelium discoideum myosin motor domain.";
RL   J. Mol. Biol. 274:394-407(1997).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC   -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC       INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC       LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC       (MLC-2).
CC   -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC       CORTEX.
CC   -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
CC       SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC       THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC       ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC       POSITION (688).
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   -!- SIMILARITY: Contains 1 IQ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M14628; AAA33227.1; -.
DR   PIR; A26655; A26655.
DR   PDB; 1MMA; 03-DEC-97.
DR   PDB; 1MMD; 17-AUG-96.
DR   PDB; 1MMG; 03-DEC-97.
DR   PDB; 1MMN; 03-DEC-97.
DR   PDB; 1MND; 17-AUG-96.
DR   PDB; 1MNE; 17-AUG-96.
DR   PDB; 1VOM; 23-DEC-96.
DR   PDB; 1LVK; 28-JAN-98.
DR   PDB; 1D0X; 20-DEC-00.
DR   PDB; 1D0Y; 20-DEC-00.
DR   PDB; 1D0Z; 20-DEC-00.
DR   PDB; 1D1A; 20-DEC-00.
DR   PDB; 1D1B; 20-DEC-00.
DR   PDB; 1D1C; 20-DEC-00.
DR   PDB; 1FMV; 20-DEC-00.
DR   PDB; 1FMW; 20-DEC-00.
DR   PDB; 1G8X; 17-JAN-01.
DR   PDB; 1JWY; 07-NOV-01.
DR   PDB; 1JX2; 07-NOV-01.
DR   DictyDb; DD01008; mhcA.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW   Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT   DOMAIN        1    761       MYOSIN HEAD-LIKE.
FT   DOMAIN      762    791       IQ.
FT   DOMAIN      817   2116       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP.
FT   DOMAIN      638    660       ACTIN-BINDING.
FT   DOMAIN      738    752       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (DI-) (POTENTIAL).
FT   MOD_RES     678    678       ALKYLATION (SH-1).
FT   MOD_RES    1823   1823       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    1833   1833       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    2029   2029       PHOSPHORYLATION (BY MHCK).
FT   TURN          3      5
FT   TURN          7      8
FT   HELIX        10     15
FT   HELIX        25     28
FT   STRAND       34     37
FT   TURN         41     42
FT   STRAND       48     55
FT   STRAND       59     63
FT   STRAND       69     73
FT   HELIX        74     76
FT   STRAND       78     79
FT   HELIX        83     85
FT   TURN         86     87
FT   STRAND       90     90
FT   HELIX        91     93
FT   HELIX        99    110
FT   TURN        111    113
FT   STRAND      116    119
FT   TURN        120    121
FT   STRAND      122    126
FT   HELIX       137    142
FT   TURN        143    145
FT   HELIX       148    150
FT   HELIX       155    169
FT   TURN        170    170
FT   STRAND      173    179
FT   TURN        181    182
FT   HELIX       185    200
FT   HELIX       210    226
FT   STRAND      227    228
FT   TURN        234    235
FT   STRAND      236    237
FT   STRAND      240    247
FT   TURN        249    250
FT   STRAND      253    261
FT   HELIX       265    268
FT   TURN        269    269
FT   TURN        273    274
FT   STRAND      278    278
FT   HELIX       279    287
FT   HELIX       290    296
FT   TURN        297    297
FT   HELIX       301    303
FT   TURN        305    307
FT   TURN        316    317
FT   HELIX       320    334
FT   TURN        335    335
FT   HELIX       338    355
FT   TURN        356    356
FT   STRAND      360    360
FT   STRAND      368    368
FT   HELIX       373    382
```

Query Match 7.8%; Score 186.5; DB 1; Length 2116;
Best Local Similarity 21.6%; Pred. No. 0.13;
Matches 108; Conservative 86; Mismatches 205; Indels 101; Gaps 17;

```
Qy         26 SQDKS--LHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE 83
              ::|||  |  ::   |     | | | |   |      ||  |  |: : |      :|
Db       1023 TKDKSELLRQKKKLEEELKQVQEALAAETAAKLAQEAANKKLQGEYTEL-NEKFNSEVTA 1081

Qy         84 KEKIEKERQSIRSS--PLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQL 141
                :  :|| :::: |    ::|:|: |      |||    |  :  |  ||   :: :| |
Db       1082 RSNVEKSKKTLESQLVAVNNELDEE----KKNR---DALEKKKKALDAMLEEMKDQLEST 1134

Qy        142 DGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLI 201
               |   :  |:       :: :|:|       :|:| :      |    ||| ||| :  :|
Db       1135 GGEKKSLYDL------KVKQESDMEALRNQISELQSTIAKLEKIKSTLEGEVARLQGELE 1188

Qy        202 SKE--ANNYEEDPNKPTSWTENQAGKIPEKVT-------------PMAAIQDGLAKGEN 245
              :::   :| |:   |      |::: :: |:                 :: :|  |::  |
Db       1189 AEQLAKSNVEKQKKKVELDLEDKSAQLAEETAAKQALDKLKKKLEQELSEVQTQLSEANN 1248

Qy        246 D-----------ETVSNTLTLTNGLERRTKTYSEDN----FEELQYFPNFYALLKSIDSE 290
                          ||  | | |    |:: |   |         ||::       : : :: |
Db       1249 KNVNSDSTNKHLETSFNNLKLELEAEQKAKQALEKKRLGLESELKH------VNEQLEEE 1302

Qy        291 KEAKEK--------ETLITIMKTLIDFVKMMVKYGTISPEEGVSYLE------------ 329
              |: ||          |  :: :|  |:       | | :  :  | |:
Db       1303 KKQKESNEKRKVDLEKEVSELKDQIEEEVASKKAVTEAKNKKESELDEIKRQYADVVSSR 1362

Qy        330 --NLDEMIALQTKN-KLEKNATDNISKLFPAPSEKSHEETD------STKEEAAKMEKEY 380
                :::::  || || :|   | :   :|  |   |   | |       : :|| ||  |
Db       1363 DKSVEQLKTLQAKNEELRNTAEEAEGQLDRAERSKKKAEFDLEEAVKNLEEETAKKVKAE 1422

Qy        381 GSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFI 440
               ::| :  |  |       |     :| |::  | | |                 ||::|  :
Db       1423 KAMKKAETDYRSTKSELDDAKNVSSEQYVQIKRLNEE--------------LSELRSVL 1467

Qy        441 NKQADAYVEKGILDKEEAEA 460
               ::||      |  |: ||:
Db       1468 -EEADERCNSAIKAKKTAES 1486
```

```
RESULT 6
USO1_YEAST
ID   USO1_YEAST     STANDARD;      PRT;  1790 AA.
AC   P25386;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Intracellular protein transport protein USO1.
GN   USO1 OR INT1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180-1A;
RX   MEDLINE=91185402; PubMed=2010462;
RA   Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA   Yamasaki M.;
RT   "A cytoskeleton-related gene, uso1, is required for intracellular
RT   protein transport in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 113:245-260(1991).
RN   [2]
RP   SEQUENCE OF 782-1790 FROM N.A.
RA   Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA   Kendrick K.E.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Bai Y., Symington L.S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC       MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC       ER AND THE GOLGI COMPLEX.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC       OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC       COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC   -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54378; CAA38253.1; -.
DR   EMBL; L03188; AAB00143.1; -.
DR   EMBL; U53668; AAB66659.1; -.
DR   SGD; S0002216; USO1.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR006955; Uso1_p115_C.
DR   InterPro; IPR006953; Uso1_p115_head.
DR   Pfam; PF04871; Uso1_p115_C; 1.
DR   Pfam; PF04869; Uso1_p115_head; 1.
KW   Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT   DOMAIN        1    724       GLOBULAR HEAD.
FT   DOMAIN      725   1790       COILED COIL (POTENTIAL).
FT   DOMAIN      465    487       CHARGED (HYPER-HYDROPHILIC).
FT   DOMAIN      991   1790       DISPENSABLE FOR THE PROTEIN FUNCTION.
FT   DOMAIN     1172   1786       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT    847    847       G -> E (IN REF. 2).
FT   CONFLICT    924    924       E -> K (IN REF. 2).
FT   CONFLICT   1253   1253       V -> I (IN REF. 2).
FT   CONFLICT   1319   1319       I -> V (IN REF. 2).
FT   CONFLICT   1461   1461       N -> S (IN REF. 2).
FT   CONFLICT   1581   1581       G -> S (IN REF. 2).
FT   CONFLICT   1600   1600       I -> V (IN REF. 2).
FT   CONFLICT   1661   1661       R -> S (IN REF. 2).
FT   CONFLICT   1772   1772       D -> DEEDDEE (IN REF. 2).

SQ   SEQUENCE   1790 AA;  206424 MW;  6CE2B216E9FD4818 CRC64;

  Query Match             7.6%;  Score 183.5;  DB 1;  Length 1790;
  Best Local Similarity   20.3%;  Pred. No. 0.15;
  Matches  120;  Conservative   91;  Mismatches  200;  Indels  179;  Gaps   23;

Qy         29 KSLHN--RELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI----T 82
              |||| |  ::: ||   || : :| |:               | | :   ||   |
Db        965 KSLANNYKDMQAE---NESLIKAVEE------------SKNESSIQLSNLQNKIDSMSQ 1008

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI------DDYDSTKSGLDHKFQD--- 133
              |||  : || ||  :    |  : |:: ||   :       |:|:|  | |  | :
Db       1009 EKENFQIERGSIEKNIEQLKKTISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATT 1068

Qy        134 -DPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH----- 187
               : : ::::     | |::  ::||  |:     :  |: :    |  : |:: |
Db       1069 ANDENVNKISELTKTREELEAELAA--YKNLKNELETKLETSEKALKEVKENEEHLKEEK 1126

Qy        188 -TLEDEVAEVLQKLISKEAN---------------NYEED-PNKPTSWTENQAGKIPEK 229
               || |  |  |:| |  ||                |||   ||   : | :  :: ::
Db       1127 IQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKERQYNE-EISQLNDE 1185

Qy        230 VTPMAAIQDGLAKGENDETVSNTLTLTNGLERRT-----------------KTYSEDN-- 270
               :|     : : | :|||       : :  | ::                 |  :| |
Db       1186 ITSTQQENESIKK-KNDELEGEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEA 1244

Qy        271 --------------FEELQYFPNF--------------------YALLKSIDSEKEAK 294
                           :|||   ||                    |  |:  :|||  :
Db       1245 SLLESIKSVESETVKIKELQDECNFKEKEVSELEDKLKASEDKNSKYLELQK-ESEKIKE 1303

Qy        295 EKETLITIMKTLIDFVKMMVK------------------------------------- 315
               | :   | :|  :: :  : |
Db       1304 ELDAKTTELKIQLEKITNLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQ 1363

Qy        316 ------------YGTISPE--EGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEK 361
                         ||: |  | :: ||  ||:| || :|:|:    ||         |
Db       1364 AFEKERKLLNEGSSTITQEYSEKINTLE--DELIRLQNENELKAKEIDNTRSELEKVSLS 1421

Qy        362 SHEETDSTKEEAAKMEKEYGSLKDS-TKDDNSNPGGKTDEP------KGKTEAYLEAIRK 414
               : |  :  :     :: |  | ||  |::|      : |         | :  |  |:  |
Db       1422 NDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAK 1481

Qy        415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILD-KEEAEAIKR 463
               | ||| ::: :||  :| | :: : |: :: :|    :  |   | |::
Db       1482 VEEGLKKLEEESSKEKAELEKSKEMM-KKLESTIESNETELKSSMETIRK 1530


RESULT 7
MAPB_RAT
ID   MAPB_RAT       STANDARD;      PRT;  2459 AA.
AC   P15205; Q62958; Q9ER21; Q9QW92;
DT   01-APR-1990 (Rel. 14, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE   light chain LC1].
GN   MAP1B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-142 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Testis;
RX   MEDLINE=96257242; PubMed=8666295;
RA   Liu D., Fischer I.;
RT   "Isolation and sequencing of the 5' end of the rat microtubule-
RT   associated protein (MAP1B)-encoding cDNA.";
RL   Gene 172:307-308(1996).
RN   [2]
RP   SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX   MEDLINE=92347374; PubMed=1639092;
```

```
RA   Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT   "Identification of two distinct microtubule binding domains on
RT   recombinant rat MAP 1B.";
RL   Eur. J. Cell Biol. 57:66-74(1992).
RN   [3]
RP   SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Spinal cord;
RX   MEDLINE=90059871; PubMed=2555150;
RA   Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA   Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT   "Neuraxin, a novel putative structural protein of the rat central
RT   nervous system that is immunologically related to microtubule-
RT   associated protein 5.";
RL   EMBO J. 8:2879-2888(1989).
RN   [4]
RP   DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX   MEDLINE=97405699; PubMed=9260743;
RA   Ma D., Nothias F., Boyne L.J., Fischer I.;
RT   "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT   in rat CNS and PNS during development.";
RL   J. Neurosci. Res. 49:319-332(1997).
CC   -!- FUNCTION: The function of brain MAPS is essentially unknown.
CC       Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC       that accompany neurite extension. Possibly MAP1B Binds to at least
CC       two tubulin subunits in the polymer, and this bridging of subunits
CC       might be involved in nucleating microtubule polymerization and in
CC       stabilizing microtubules.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC       cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC       heart or muscle.
CC   -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC       nerve levels are high early in development but decrease during
CC       postnatal development and are low in adults. In dorsal root
CC       ganglia levels remain high throughout development.
CC   -!- INDUCTION: By nerve growth factor.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC       from MAP1B by proteolytic processing. It is free to associate with
CC       both MAP1A and MAP1B. It interacts with the amino-terminal region
CC       of MAP1B (By similarity).
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: TO MAP1A.
CC   -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC       2459) was originally described as neuraxin in Ref.3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U52950; AAB17068.1; -.
DR   EMBL; X60370; CAC16162.1; -.
DR   EMBL; X16623; CAA34620.1; ALT_SEQ.
DR   PIR; A56577; A56577.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 8.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?    2459       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1869    1885       MAP1B 1.
FT   REPEAT     1886    1902       MAP1B 2.
FT   REPEAT     1903    1919       MAP1B 3.
FT   REPEAT     1920    1936       MAP1B 4.
FT   REPEAT     1937    1953       MAP1B 5.
FT   REPEAT     1954    1970       MAP1B 6.
FT   REPEAT     1988    2004       MAP1B 7.
FT   REPEAT     2005    2021       MAP1B 8.
FT   REPEAT     2022    2038       MAP1B 9.
FT   REPEAT     2039    2055       MAP1B 10.
FT   DOMAIN      559    1035       GLU-RICH.
FT   DOMAIN      588     786       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                 KKEE AND KKEI/V REPEATS).
FT   DOMAIN     2224    2312       LYS-RICH.
FT   CONFLICT    127     127       M -> V (IN REF. 1).
FT   CONFLICT    140     140       T -> S (IN REF. 1).
FT   CONFLICT   2112    2112       R -> K (IN REF. 3).
FT   CONFLICT   2169    2169       L -> I (IN REF. 3).
SQ   SEQUENCE   2459 AA;  269497 MW;  2E3F6872DEDB8BA2 CRC64;

  Query Match             7.5%;  Score 180.5;  DB 1;  Length 2459;
  Best Local Similarity   22.7%;  Pred. No. 0.31;
  Matches  120;  Conservative   77;  Mismatches  200;  Indels  131;  Gaps   22;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIK-KTYPPENKPGQSNYSFVDNLNLLKA 80
              ||  |:     ::| : |      |: :  : : | |    ::|||:        :    :
Db        549 KPLSSKSVRKESKEEAPEATKASQVEKTPKVESKEKVIVKKDKPGK--------VESKPS 600

Qy         81 ITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDP--DGL 138
              :||||   || ||   : :  |   |      |::     ||   : | :: |  :
Db        601 VTEKEVPSKEEQSPVKAEVAEKAATESKPKVTKDKVVKKEIKTKP--EEKKEEKPKKEVA 658

Qy        139 HQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHT-LEDEVAEVL 197
               : | |||  ::   |       |  :    | : |     |  | :  | |:|   ||
Db        659 KKEDKTPLKKDEKPKK-------EEAKKEIKKEIKKEEKKELKKEVKKETPLKDAKKEV- 710

Qy        198 QKLISKEANNYEEDPNK------------PTSWTENQAG--------KIPEKVTPMAA- 235
              :|   ||    |::| |              | | |:  |          : | |  |:||
Db        711 KKDEKKEVKKEEKEPKKEIKKISKDIKKSTPLSDTKKPAALKPKVAKKEEPTKKEPIAAG 770

Qy        236 -IQD-GLAK-----GENDETVSNTL----------------------TLTNGLERRTK 264
              ::| |  |      |:  |  :  :                       :| :  |  ||
Db        771 KLKDKGKVKVIKKEGKTTEAAATAVGTAAVAAAAGVAASGPAKELEAERSLMSSPEDLTK 830

Qy        265 TYSEDNFEELQYFPNFYALLKSIDSEKEAKE----------KETLITIMKTLIDFVKMMV 314
              : |   ||:    :    |: |: |:: ||          ||| ::
Db        831 DFEELKAEEIDVAKDIKPQLELIEDEEKLKETEPGEAYVIQKETEVS------------ 877

Qy        315 KYGTISPEEGVSYLENLDEMIALQTKNKL---EKNATDNISKL------FPAPSEKSHEE 365
              |    ||:||::  |    |    ||  :|  ||   |:| |       |   ||    |
Db        878 KGSAESPDEGITTTEGEGE--CEQTPEELEPVEKQGVDDIEKFEDEGAGFEESSEAGDYE 935

Qy        366 TDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEA--YLEAIRKNI------- 416
              :  ||| : |::    :|:     |      ||   | ||  :::  |:::
Db        936 EKAETEEAEEPEEDG---EDNVSGSASKHSPTEDEEIAKAEADVHIKEKRESVASGDDRA 992

Qy        417 -----EWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE 459
                  | |:| : : ::|: :  :     :|  ||  |    || |||
Db        993 EEDMDEALEKGEAEQSEEEGEEEE-----DKAEDAREEDHEPDKTEAE 1035


RESULT 8
RA50_METJA
ID   RA50_METJA     STANDARD;      PRT;  1005 AA.
AC   Q58718;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR MJ1322.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
```

```
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67572; AAB99331.1; -.
DR   PIR; A64465; A64465.
DR   TIGR; MJ1322; -.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003405; SMC_C.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF04423; Rad50_zn_hook; 1.
DR   Pfam; PF02483; SMC_C; 1.
DR   Pfam; PF02463; SMC_N; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      32     39       ATP (BY SIMILARITY).
FT   DOMAIN      158    849       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1005 AA;  119387 MW;  9BBBB48173E788F3 CRC64;

  Query Match             7.5%;  Score 179;  DB 1;  Length 1005;
  Best Local Similarity   22.3%;  Pred. No. 0.12;
  Matches  116;  Conservative   80;  Mismatches  164;  Indels  160;  Gaps   27;

Qy         52 DKIKKTYPPENKPGQSNYSFVDNLNLLKA-ITEKEKIEKERQSIRSSPLDNKLN-VEDVD 109
              |: :| |    | |:    :   |  ::  :  ||  |||        | ||:: :|:
Db        159 DEFEKCY---QKMGEIVKEYEKRLERIEGELNYKENYEKE--------LKNKMSQLEE-- 205

Qy        110 STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFD 169
                ||:||::  |     :  :|:|     :: :   |  |  ::|:     ||  ||:
Db        206 --KNKKLMEINDKLNK-IKKEFEDIEKLFNEWENKKLLYEKFINKL-----EERKRALEL 257

Qy        170 KIVS-KLLNLGLITESQAH-------------TLEDEVAEVLQKLISKEANNYEEDPNK 214
              |    |:|   | |  :|               :| ||: ::  :|  :|  :: ||  |
Db        258 KNQELKILEYDLNTVVEARETLNRHKDEYEKYKSLVDEIRKIESRL--RELKSHYEDYLK 315

Qy        215 PTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEEL 274
              |  |   | | ||      ::: : | :  : : |  || | :        :|  | :
Db        316 LTKQLEIIKGDI-EK------LKEFINKSKYRDDIDNLDTLLNKI--------KDEIERV 360

Qy        275 QYFPNFYALLKSIDSEKEAKEK------------ETLITIMKTLIDFVKMMVKYGTISPE 322
              :   :    ||::: | |  ||            |  : : :  ::: |: ::| |:
Db        361 ETIKDLLEELKNLNEEIEKIEKYKRICEECKEYYEKYLELEEKAVEYNKLTLEYITL--- 417

Qy        323 EGVSYLENLDEMIALQTKNKLEKNATD---NISKLFPAPS-------EKSHEETDSTK-- 370
                        || |  :|||  |    |:||             | | :| :  |
Db        418 -------------LQEKKSIEKNINDLETRINKLLEETKNIDIESIENSLKEIEEKKKV 463

Qy        371 -----EEAAKMEKEYGSLKDSTK------DDNSNPGGK-------TDEPK-------GKT 405
```

```
                :|  :: |: | :    |      |:      ||        || |        ||
Db        464 LENLQKEKIELNKKLGEINSEIKRLKKILDELKEVEGKCPLCKTPIDENKKMELINQHKT 523

Qy        406 E---AYLE---------AIRKNIEWLKKH--------------DKKGNKEDYDL----- 433
              :    | |           | |:|| |||                :|:   |: :|
Db        524 QLNNKYTELEEINKKIREIEKDIEKLKKEIDKEENLKTLKTLYLEKQSQIEELELKLKNY 583

Qy        434 SKMRDFINKQADAYVEKG-----ILD--KEEAEAIKRIYS 466
               :  | |||:   ||  |     ||:  | :    |  |:
Db        584 KEQLDEINKKISNYVINGKPVDEILEDIKSQLNKFKNFYN 623


RESULT 9
REST_HUMAN
ID   REST_HUMAN     STANDARD;      PRT;  1427 AA.
AC   P30622;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
DE   Sternberg intermediate filament associated protein).
GN   RSN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood monocytes;
RX   MEDLINE=92289675; PubMed=1600942;
RA   Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M.,
RA   Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
RA   de Wolf-Peeters C., Shipman R.;
RT   "Restin: a novel intermediate filament-associated protein highly
RT   expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL   EMBO J. 11:2103-2113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92405160; PubMed=1356075;
RA   Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT   "CLIP-170 links endocytic vesicles to microtubules.";
RL   Cell 70:887-900(1992).
CC   -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC       THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC       CYTOSKELETON.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P30622-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P30622-2; Sequence=VSP_000765;
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC       OF HODGKIN'S DISEASE.
CC   -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64838; CAA46050.1; -.
DR   EMBL; M97501; AAA35693.1; -.
DR   PIR; S22695; S22695.
DR   Genew; HGNC:10461; RSN.
DR   MIM; 179838; -.
DR   GO; GO:0005768; C:endosome; TAS.
DR   GO; GO:0005882; C:intermediate filament; TAS.
DR   GO; GO:0015630; C:microtubule cytoskeleton; TAS.
DR   GO; GO:0008017; F:microtubule binding activity; TAS.
```

```
DR   GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR   InterPro; IPR000938; CAP-Gly.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01302; CAP_GLY; 2.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 2.
DR   PROSITE; PS50245; CAP_GLY_2; 2.
KW   Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
FT   DOMAIN       78    120       CAP-GLY 1.
FT   DOMAIN      143    204       SER-RICH.
FT   DOMAIN      232    274       CAP-GLY 2.
FT   DOMAIN      304    331       SER-RICH.
FT   DOMAIN      350   1342       COILED COIL (POTENTIAL).
FT   DOMAIN     1408   1421       CCHC-BOX.
FT   VARSPLIC    457    491       Missing (in isoform Short).
FT                                /FTId=VSP_000765.
FT   CONFLICT   1069   1069       D -> E (IN REF. 2).
SQ   SEQUENCE   1427 AA;  160989 MW;  0A4F166DD94254E8 CRC64;

  Query Match             7.4%;  Score 178.5;  DB 1;  Length 1427;
  Best Local Similarity   22.4%;  Pred. No. 0.2;
  Matches  111;  Conservative   79;  Mismatches  182;  Indels  123;  Gaps   20;

Qy         34 RELSAE-----RPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL--------NLLKA 80
              |: |:|     : | :|: || | :||     :|       |     |         || :
Db        769 RKASSEGKSEMKKLRQQL-EAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTNLQEN 827

Qy         81 ITE----KEKIEKERQ-------------SIRSSPLD--NKL-------NVEDVDSTKN 113
              ::|    || :||| |               |:: |  :  |||       |:    | |
Db        828 LSEVSQVKETLEKELQILKEKFAEASEEAVSVQRSMQETVNKLHQKEEQFNMLSSDLEKL 887

Qy        114 RKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVS 173
              |: : |       :: ||::  :   ||              | |:   |||       ::
Db        888 RENLAD-------MEAKFREKDEREEQL-------------IKAKEKLEND-------IA 920

Qy        174 KLLNLGLITESQAHTLEDEV------AEVLQKLISKEANNYE------EDPNKPTSWTEN 221
              ::: :     ||   : ||:       | ||  ::|   |         ||       ::
Db        921 EIMKMSGDNSSQLTKMNDELRLKERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQ 980

Qy        222 QAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFY 281
              :| |  |:           | | :  :|:       ||:: :| | :  :||:
Db        981 EAAKKHEE-----------EKKELERKLSD-------LEKKMET-SHNQCQELK------ 1015

Qy        282 ALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKN 341
              |  :   || : | :| |  : |||:|    :      : ||    |: |:|:    |
Db       1016 ARYERATSETKTKHEEILQNLQKTLLDTEDKLKG----AREENSGLLQELEELRKQADKA 1071

Qy        342 KLEKNATDNISKLFPAPSEKSH--EETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTD 399
              |  : | | :  :     ||:      : ||:  ||:| |  :||::     |      |:
Db       1072 KAAQTAEDAMQIMEQMTKEKTETLASLEDTKQTNAKLQNELDTLKENNL-KNVEELNKSK 1130

Qy        400 EPKGKTEAYLEAIRKNIEWLKKH----------DKKGNKEDYDLSKMRDFINKQADAYV 448
              |         :|  || || ||:           ::  |   :| : || :
Db       1131 ELLTVENQKMEEFRKEIETLKQAAAQKSQQLSALQEENVKLAEELGRSRDEVTSHQKLEE 1190

Qy        449 EKGILDKEEAEAIKR 463
              |: :|: :  |  ||
Db       1191 ERSVLNNQLLEMKKR 1205


RESULT 10
RA50_PYRFU
ID   RA50_PYRFU     STANDARD;      PRT;   882 AA.
AC   P58301;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR PF1167.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX   MEDLINE=20485553; PubMed=11029422;
RA   Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
RA   Carney J.P.;
RT   "Mre11 and Rad50 from Pyrococcus furiosus: cloning and biochemical
RT   characterization reveal an evolutionarily conserved multiprotein
RT   machine.";
RL   J. Bacteriol. 182:6036-6041(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX   MEDLINE=20348838; PubMed=10892749;
RA   Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
RA   Carney J.P., Tainer J.A.;
RT   "Structural biology of Rad50 ATPase: ATP-driven conformational
RT   control in DNA double-strand break repair and the ABC-ATPase
RT   superfamily.";
RL   Cell 101:789-800(2000).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site.
CC   -!- SUBUNIT: Forms a complex with mre11.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010225; AAL81291.1; -.
DR   PDB; 1F2T; 20-SEP-00.
DR   PDB; 1F2U; 02-AUG-00.
DR   PDB; 1II8; 30-MAY-01.
DR   PDB; 1L8D; 28-AUG-02.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF04423; Rad50_zn_hook; 1.
DR   Pfam; PF02463; SMC_N; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
KW   Complete proteome.
FT   NP_BIND      30     37       ATP.
FT   DOMAIN      148    744       COILED COIL (POTENTIAL).
SQ   SEQUENCE   882 AA;  103839 MW;  3ADCBD250382A99E CRC64;

  Query Match             7.4%;  Score 178;  DB 1;  Length 882;
  Best Local Similarity   22.1%;  Pred. No. 0.12;
  Matches  112;  Conservative   90;  Mismatches  172;  Indels  132;  Gaps   21;

Qy         39 ERPLNEQIAEAEE--DKIK---KTYPPENKPGQSNYSFVDNLNLLKA-----ITEKEKIE 88
              :| | |:|   ||  :|:|   |    : |   |    ||     ||       : :| |||
Db        260 KRTLEERIKNTEEYLEKLKEKEKELEEQVKEITSIKKDVDAYLALKEFKNEYLDKKYKIE 319

Qy         89 KERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTA 148
              ||    |   | |::        |  : ::: :| |  |:::
Db        320 KE--LTRVEELINEIQ-------KRIEELNEKESEKEKLENE----------------K 353

Qy        149 EDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNY 208
              ::|::|:|     | |  ::::| :|  ||  : |       ::: ::|::| :|:
```

```
Db        354 KEILNKLAIL---EKDHQLYEEIKAKKENLRQLKEKLGDKSPEDIKKLLEELETKK-TTI 409

Qy        209 EEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL--AKGE--------NDE----------- 247
Db        410 EEERNEIT----QRIGELKNKIGDLKTAIEELKKAKGKCPVCGRELTDEHREELLSKYHL 465

Qy        248 -------TVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDS----------E 290
Db        466 DLNNSKNTLAKLIDRKSELERELRRIDME-IKRLTPLLTVAEQIRSIEEELNVVNLEKIE 524

Qy        291 KEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIA--LQTKNKLEKNAT 348
Db        525 KNATEYEKLLEELRTLEGRIRGLA--------EDLKKLAPLEKKLAALIHKKQELEKELK 576

Qy        349 DNISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKD---------DNSNPGGKTD 399
Db        577 ELNTKL-ESFGFKSVEDLDSKLRELEEIYKRYLTLLNSKKELEITQREIAKAKETLEMSF 635

Qy        400 EPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDY----------------------DLSKMR 437
Db        636 EELAEVEADIERIEKKLSQLK---QKYNEEEYKKKREEKEELEKELARLEAQKKELEKRR 692

Qy        438 DFINKQADAYVEKGILDKEEAEAIKR 463
Db        693 DTIK----STLEKLKAEKENRERVKK 714
```

```
RESULT 11
BAG_STRAG
ID   BAG_STRAG      STANDARD;      PRT;  1164 AA.
AC   P27951;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   IgA FC receptor precursor (Beta antigen) (B antigen).
GN   BAG.
OS   Streptococcus agalactiae.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1311;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
RC   STRAIN=LA239;
RX   MEDLINE=91312121; PubMed=1857207;
RA   Jerlstroem P.G., Chhatwal G.S., Timmis K.N.;
RT   "The IgA-binding beta antigen of the c protein complex of Group B
RT   streptococci: sequence determination of its gene and detection of two
RT   binding regions.";
RL   Mol. Microbiol. 5:843-849(1991).
RN   [2]
RP   IDENTIFICATION OF IG-LIKE DOMAIN.
RX   MEDLINE=97035265; PubMed=8880921;
RA   Bateman A., Eddy S.R., Chothia C.;
RT   "Members of the immunoglobulin superfamily in bacteria.";
RL   Protein Sci. 5:1939-1942(1996).
CC   -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC       an amide bond (Potential).
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59771; CAA42442.1; -.
DR   PIR; S15330; FCSOAG.
DR   InterPro; IPR004829; Csurface_antigen.
DR   InterPro; IPR005877; Gpos_YSIRK.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003599; Ig.
```

```
DR   InterPro; IPR006192; LPXTG.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF05062; RICH; 1.
DR   Pfam; PF04650; YSIRK_signal; 1.
DR   ProDom; PD153432; Csurface_antigen; 1.
DR   SMART; SM00409; IG; 1.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR   PROSITE; PS50835; IG_LIKE; FALSE_NEG.
KW   Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
KW   Immunoglobulin domain.
FT   SIGNAL        1     37
FT   CHAIN        38   1135       IGA FC RECEPTOR.
FT   PROPEP     1136   1164       REMOVED BY SORTASE (POTENTIAL).
FT   DOMAIN      434    534       IG-LIKE.
FT   DOMAIN      199    438       IGA-BINDING (POTENTIAL).
FT   DOMAIN      439    826       IGA-BINDING (POTENTIAL).
FT   DOMAIN      827    945       PRO-RICH REPEATS.
FT   SITE       1132   1136       LPXTG SORTING SIGNAL (POTENTIAL).
FT   MOD_RES    1135   1135       AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ   SEQUENCE   1164 AA;  131051 MW;  65DE94AF720A5474 CRC64;

  Query Match             7.2%;  Score 172.5;  DB 1;  Length 1164;
  Best Local Similarity   21.6%;  Pred. No. 0.3;
  Matches  124;  Conservative   77;  Mismatches  189;  Indels  185;  Gaps   25;

Qy         36 LSAERPLNE--------QIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLLKAIT-EKEK 86
Db        149 LELENQFNETNRLLHIKQHEEVEKDKKAK----QQKTLKQSDTKVDLSNIDKELNHQKSQ 204

Qy         87 IEK--ERQSIRSSPLDNKL--------------NVEDVDSTKNRKLIDDYDSTKSGLDHK 130
Db        205 VEKMAEQKGITNEDKDSMLKKIEDIRKQAQQADKKEDAEVKVREELGKLFSSTKAGLDQE 264

Qy        131 FQDDPDGLHQLDGTPLTAEDIVHKI-----------AARIYEENDRAVFDKIVSKLLNLG 179
Db        265 IQE-----HVKKET--SSEENTQKVDEHYANSLQNLAQKSLEELDKATTNEQATQVKNQF 317

Qy        180 L-------------------ITESQAHTLEDEVAEV-------LQKLISKE---ANNYEE 210
Db        318 LENAQKLKEIQPLIKETNVKLYKAMSESLEQVEKELKHNSEANLEDLVAKSKEIVREYEG 377

Qy        211 DPNKPTSWTE-------------------NQAGKIPEKVTPMAAIQDGLAKGENDETVSN 251
Db        378 KLNQSKNLPELKQLEEEAHSKLKQVVEDFRKKFKTSEQVTPKKRVKRDLAANENNQ---Q 434

Qy        252 TLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVK 311
Db        435 KIELTVSPENITVYEGED--------VKFTVTAKS-DS-------------KTTLDFSD 471

Qy        312 MMVKYG-TISPEEGVSYLENLD-----------------EMIALQTK----NKLEKNATD 349
Db        472 LLTKYNPSVSDRISTNYKTNTDNHKIAEITIKNLKLNESQTVTLKAKDDSGNVVEKTFTI 531

Qy        350 NISKLFPAPSEKSHEETDSTKEEAA-----------------------KMEKEYGSLKD 385
Db        532 TVQKKEEKQVPKTPEQKDSKTEEKVPQEPKSNDKNQLQELIKSAQQELEKLEKAIKELME 591

Qy        386 STKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKK----HDKKGNKEDYDLSKMRDFIN 441
Db        592 Q-PEIPSNP--EYGIQKSIWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMN 648

Qy        442 KQADAYVEKGILDKEEAE---------AIKRIYSS 467
Db        649 YQLHAQME--MLTRKVVQYMNKYPDNAEIKKIPES 681
```

```
RESULT 12
TRDN_CANFA
ID   TRDN_CANFA     STANDARD;      PRT;   700 AA.
AC   P82179;
DT   28-FEB-2003 (Rel. 41, Created)
```

```
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Triadin.
GN   TRDN.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Heart, and Skeletal muscle;
RX   MEDLINE=99428545; PubMed=10497235;
RA   Kobayashi Y.M., Jones L.R.;
RT   "Identification of triadin 1 as the predominant triadin isoform
RT   expressed in mammalian myocardium.";
RL   J. Biol. Chem. 274:28660-28668(1999).
CC   -!- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRIN TO THE
CC       JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC       COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC       reticulum.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=Skeletal;
CC         IsoId=P82179-1; Sequence=Displayed;
CC       Name=Cardiac 1;
CC         IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
CC       Name=Cardiac 3;
CC         IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
CC   -!- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF165916; AAF00222.1; -.
DR   EMBL; AF165915; AAF00221.1; -.
DR   EMBL; AF165917; AAF00223.1; -.
KW   Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW   Alternative splicing.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     47     67       POTENTIAL.
FT   DOMAIN       68    700       LUMENAL (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    257    277       DQYAFCRYMIDMFVHGDLRPG -> GKHSEEVAGGSKRTLG
FT                                KKQIQ (in isoform Cardiac 1).
FT                                /FTId=VSP_004001.
FT   VARSPLIC    278    700       Missing (in isoform Cardiac 1).
FT                                /FTId=VSP_004002.
FT   VARSPLIC    466    466       E -> EPIKGKEVKVPGSLKEKE (in isoform
FT                                Cardiac 3).
FT                                /FTId=VSP_004003.
FT   VARSPLIC    524    579       EEKVVKQVKATEKAAIEKTVKPKPAKKAEHQEKESPTIKTD
FT                                KPKPTSKETPEVTES -> GILQVVPVVLNCLFLVQFQQDE
FT                                ELNVESKVFRMIHVLSHPTSRTSPILVISTTCRT (in
FT                                isoform Cardiac 3).
FT                                /FTId=VSP_004004.
SQ   SEQUENCE   700 AA;  78152 MW;  F033E3AA1BEE0C56 CRC64;

  Query Match             7.0%;  Score 169;  DB 1;  Length 700;
  Best Local Similarity  20.0%;  Pred. No. 0.24;
  Matches   97;  Conservative   75;  Mismatches  176;  Indels  136;  Gaps   19;

Qy         27 QDKSLH----NRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT 82
              : | :|     :|   || :  ::|  |::|:|:    | |                 |
Db        139 KQKEIHKEKAEKEEKPERKILAKVAHREKEKVKEKEKSEKK----------------AT 181

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLD 142
              ||||||: :       : |   ::    |  : |       | |   | :     : ::
Db        182 HKEKIEKKEKP------ETKTMAKEERKAKTEEKIK--KEVKGGKQEKVKPTAAKVKEVQ 233

Qy        143 GTPLTAEDIVHKIAARI--YEENDRAVFDKIVSKLLNLGLITESQAHTL---------- 189
              ||  |::   |  | :  :|: |:  | : :  :   | :   |:  |
Db        234 KTPPKAKEKEGKETAAVAKHEQKDQYAFCRYMIDMFVHGDLRPGQSPALPPPLPTVQASR 293

Qy        190 ----------EDEVAEVLQKLISKEANNYEEDPNKPTS-------WTENQAGKIPE--- 228
                       |:|  :  :| :: |    |::  |  |         : : || ||
Db        294 PTPASPTLEGKEEEEKKKAEKKVTSETKKKEKEDVKKKSDKDTAIDVEKKEPGKAPETKQ 353

Qy        229 ---KVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLK 285
                 ||   || :   | | |    | |    :|   |   ::  |                |
Db        354 GTIKVVAQAAAK----KDEKKEDSKKTKT---PVEEHPKGKKQEKKE----------K 394

Qy        286 SIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEK 345
               ::  | :|:: :                    :| |         : :  :|:   |:
Db        395 YVEPAKSSKKEHS--------------------APSE---------KQVKAKTERAKEE 424

Qy        346 NATDNISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDE--PKG 403
               :  :  | |    |  |:|  | |:  : ||   | | ||   :  |  | | ||  ||
Db        425 TSAASTKKAVPG---KKEEKTTKTVEQEIRKEK---SGKTSTASKDKEPEIKKDEKMPKA 478

Qy        404 KTEAYLEAIRKNIEWLKKHD---KKGNKEDYDLSKMRDFIN---------KQADAYVEKG 451
                 |   :  :  ::  :| :   ||  | : |::|   ::          ||  |  ||
Db        479 DKEVKPKPPQSQVKKEEKSESQVKKEAKPEQDIAKPEKTVSHGKPEEKVVKQVKA-TEKA 537

Qy        452 ILDK 455
               ::|
Db        538 AIEK 541


RESULT 13
KF5C_HUMAN
ID   KF5C_HUMAN     STANDARD;      PRT;   957 AA.
AC   O60282; O95079;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
DE   2).
GN   KIF5C OR NKHC2 OR KIAA0531.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [2]
RP   SEQUENCE OF 355-585 FROM N.A.
RA   Engelender S., Sharp A.H., Colomer V., Tokito M.K., Lanahan A.,
RA   Worley P., Holzbaur E.L.F., Ross C.A.;
RT   "Huntingtin associated protein 1 (HAP1) interacts with the p150Glued
RT   subunit of dynactin.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, PROSTATE, AND
CC       TESTIS, AND MODERATE EXPRESSION IN KIDNEY, SMALL INTESTINE, AND
CC       OVARY.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
```

```
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB011103; BAA25457.1; -.
DR   EMBL; AF010146; AAD01436.1; -.
DR   HSSP; P56536; 2KIN.
DR   Genew; HGNC:6325; KIF5C.
DR   MIM; 604593; -.
DR   GO; GO:0005871; C:kinesin complex; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    386       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      406    923       COILED COIL.
FT   DOMAIN      859    956       GLOBULAR.
FT   DOMAIN      174    315       MICROTUBULE-BINDING.
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
FT   CONFLICT    355    360       TLKNVI -> STHASV (IN REF. 2).
FT   CONFLICT    583    585       EFT -> DRV (IN REF. 2).
SQ   SEQUENCE   957 AA;  109494 MW;  A9F25BB1C994322A CRC64;

  Query Match             7.0%;  Score 167.5;  DB 1;  Length 957;
  Best Local Similarity   20.3%;  Pred. No. 0.42;
  Matches  103;  Conservative   98;  Mismatches  182;  Indels  125;  Gaps   23;

Qy         25 GSQDKSLHNR-----ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLL- 78
              | : |:: |       ||:||             : || |  | :  ::  : : :| :
Db        321 GQRAKTIKNTVSVNLELTAE------------EWKKKYEKEKEKNKTLKNVIQHLEMEL 367

Qy         79 ------KAITEKEKIEKERQSIRSSPLDNKLNVEDVD------STKNRKLIDDYDSTKSG 126
                  :|: | |:|  : |      | ||   ::::        ||: :  : ||   |
Db        368 NRWRNGEAVPEDEQISAKDQK-NLEPCDNTPIIDNIAPVVAGISTEEK---EKYDEEISS 423

Qy        127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEND-----RAVFDKIVSKLLNLGLI 181
              |  :  |  | ::|           :  |:  :: ::::      |  ::||  :|  | :
Db        424 LYRQLDDKDDEINQ-------QSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIE 476

Qy        182 TESQAHTLEDEVAEVLQKLISKEANNYE---EDPNKPTSWTENQAGKIPEKVTPMAAIQD 238
               |:     :||| |||| |  :  | ||:   ::    |   |    :: :| | :   |
Db        477 NEA----AKDEVKEVLQAL-EELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQR 531

Qy        239 GLAK-----------------------------GENDETVSNTLTLTNGLERR----TK 264
              |::                            | ||     ||   ||:         :
Db        532 ELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTND---VKTLADVNGVIEEEFTMAR 588

Qy        265 TYSEDNFEELQYFPNFYALLKS--IDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPE 322
              |       |::   |    |:|  :|| ::    |  :   : ||   :  :|  |
Db        589 LYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLT---- 644

Qy        323 EGVSYLENLDEMIALQTKNKLEK---NATDNISKLFPAPSEKSHEETDSTKE-------- 371
                |::|::      | : :||:   : :: ::||     || || :   ||
Db        645 ---DYMQNME-----QKRRQLEESQDSLSEELAKL--RAQEKMHEVSFQDKEKEHLTRLQ 694

Qy        372 EAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDY 431
```

```
              :| :|:|      :| :: :    :  :   : :  :: ||   : |:   :|    ||
Db        695 DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEK-LSSDY 753

Qy        432 DLSKMRDFINKQADAYVEKGIL--DKEE 457
                :  |: |   :: :  :|| :|  || |
Db        754 NKLKIED---QEREMKLEKLLLLNDKRE 778


RESULT 14
MLP1_YEAST
ID   MLP1_YEAST     STANDARD;      PRT;  1875 AA.
AC   Q02455;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin-like protein MLP1.
GN   MLP1 OR YKR095W OR YKR415.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=93247549; PubMed=8483450;
RA   Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT   "A new yeast gene with a myosin-like heptad repeat structure.";
RL   Mol. Gen. Genet. 237:359-369(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC       REPAIR.
CC   -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC   -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L01992; AAA34783.1; -.
DR   EMBL; X73541; CAA51948.1; -.
DR   EMBL; Z28320; CAA82174.1; -.
DR   PIR; S38173; S38173.
DR   SGD; S0001803; MLP1.
DR   GO; GO:0005635; C:nuclear membrane; IDA.
DR   GO; GO:0005654; C:nucleoplasm; IDA.
DR   GO; GO:0006606; P:protein-nucleus import; IDA.
KW   Coiled coil; DNA repair.
FT   DOMAIN       69    487       COILED COIL (POTENTIAL).
FT   DOMAIN      531   1678       COILED COIL (POTENTIAL).
FT   DOMAIN     1834   1866       COILED COIL (POTENTIAL).
FT   CONFLICT    301    301       R -> A (IN REF. 1).
SQ   SEQUENCE   1875 AA;  218455 MW;  683A0D34C9066867 CRC64;

  Query Match             6.9%;  Score 166.5;  DB 1;  Length 1875;
  Best Local Similarity   23.6%;  Pred. No. 1;
  Matches  112;  Conservative   85;  Mismatches  178;  Indels   99;  Gaps   28;

Qy         30 SLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITE-KEKIE 88
                :| : ::   |  : |   | ||  :| : | |:    ::|| |     :| :|| |  :|
Db        588 TLKSEKMDLESRIEELQKELEE--LKTSVPNED----ASYSNV----TIKQLTETKRDLE 637
```

```
Qy         89 KERQSIRSSPLDNKLNVEDVDSTKNRKLI-----DDYDSTKSGLDHKFQDDPDGLHQLDG 143
              : |     |  :::    :||:|  |:     | ||| || :  |      | :
Db        638 SQVQD-----LQTRISQITRESTENMSLLNKEIQDLYDS-KSDISIK-------LGKEKS 684

Qy        144 TPLTAED---IVHKIAARIYEEND--RAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQ 198
              : : ||:   ::         |||  |  || : : :|     :|: |   :|
Db        685 SRILAEERFKLLSNTLDLTKAENDQLRKRFDYLQNTILK----QDSKTHETLNEYVSCKS 740

Qy        199 KL--ISKEANNYEEDPNKPTSWTEN---QAGKI-PEK------VTPMAAIQ----DGLAK 242
              ||  :  |  | :|:        :|   :  |: |||      || :  :|    | | :
Db        741 KLSIVETELLNLKEEQKLRVHLEKNLKQELNKLSPEKDSLRIMVTQLQTLQKEREDLLEE 800

Qy        243 ---------GENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEA 293
                      | :: :|     |:  :   |   |||   :::: |    ||
Db        801 TRKSCQKKIDELEDALSELKKETSQKDHHIKQLEEDNNSNIEWYQNKIEALK-------- 852

Qy        294 KEKETLIT-IMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNIS 352
              |: |::|| :     |  |:  |  ::  |     :|  :: | | | | :::   |:
Db        853 KDYESVITSVDSKQTDIEKLQYKVKSLEKE-----IE--EDKIRLHTYNVMDETINDDSL 905

Qy        353 KLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAI 412
              :       |  |::     :|    |||  | ::|         | ||     : :   |
Db        906 R-------KELEKSKINLTDAYSQIKEYKDLYETTSQSLQQTNSKLDE---SFKDFTNQI 955

Qy        413 RKNIEWLKKHDKKGNKED-YDLSKMRDF-INKQADAYVEKGILDKEEAEAIKRI 464
              ||:      |:| : ||   | | : | :| : |  ::|  ::||:|:  |||
Db        956 -KNLT-----DEKTSLEDKISLLKEQMFNLNNELD--LQKKGMEKEKADFKKRI 1001
```

RESULT 15
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC   KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC   OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC   AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC   CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC   QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; Z15005; CAA78727.1; -.
DR PIR; S28261; S28261.
DR HSSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; -.
DR MIM; 117143; -.
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISc; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN   1   335   KINESIN-MOTOR.
FT DOMAIN   336   2471   COILED COIL (POTENTIAL).
FT DOMAIN   2472   2663   GLOBULAR (POTENTIAL).
FT NP_BIND   86   93   ATP (BY SIMILARITY).
SQ SEQUENCE   2663 AA; 312087 MW; CEFC13880C8C8CB8 CRC64;

  Query Match   6.9%; Score 166; DB 1; Length 2663;
  Best Local Similarity   21.8%; Pred. No. 1.7;
  Matches 119; Conservative 103; Mismatches 175; Indels 150; Gaps   28;

```
Qy         26 SQDKSLHNRELSAERPLN---EQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAIT 82
              :|:|   |  :  |  |:   :::|| ||   :|:   :  |  |         ||: : ::
Db       1102 AQEK---NHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQL-------LNVQEEMS 1151

Qy         83 EKEKIEKERQSIRSSPLDNKLNVEDVDSTK---NRKLIDDYDSTKS---------GLDHK 130
              | :|   | :::::   : :| :| ::: :    :|| ::|:  ||          |
Db       1152 EMQKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVLKELQKS 1211

Qy        131 FQDDPDGL----HQLDGTPL-TAEDI----VH-KIAARIYEENDRAVFDKIVSKLLNLGL 180
              |: : | |     ::: | | | |::    :| |       :|  |:| :|  ::::|
Db       1212 FETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEK-TAQIINTQD 1270

Qy        181 ITESQAHTLEDEVAEVL---QKLIS--KEANNYEEDPNKPTSWTENQAGKIPEKVTPMAA 235
              : :|  ||   |   ||   |:|:   |: :  :|  |:   ||    |     | :|
Db       1271 LEKS--HTKLQEEIPVLHEEQELLPNVKKVSETQETMNELELLTEQSTTK---DSTTLAR 1325

Qy        236 IQ----------------DGLAK-GENDETVSNTL------------TLTNGLERRTK 264
              |:                | |  :| :|:   |              ||    | ::|
Db       1326 IEMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSK 1385

Qy        265 TYSEDNFEEL-----------QYFPNFYALLK-------------SIDSEKE-AKEKE 297
                  | :|             |: |   |||:             | |  |  ||||:
Db       1386 QEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKD 1445

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIA--LQTKNKLE------KNATD 349
              |  : : |           |    ||: |::|  |:|: :|:     |   :
Db       1446 DLQRLQEVL-------------QSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEE 1491

Qy        350 NISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTE--- 406
              |::|   |||         : | : ::|:  :: |  ::        |  |   | |
Db       1492 TINELRVNLSEK--------ETEISTIQKQLEAINDKLQN-------KIQEIYEKEEQLN 1536

Qy        407 -AYLEAIRKNIEWLKKHDKKGNKEDYDL----SKMRDFINKQADAYVEKGILDKEEAEAI 461
               :  :::|:  ||:  :    :|  |    ||| :  |:  ::  |  |: ||: | :
Db       1537 IKQISEVQENVNELKQFKEHRKAKDSALQSIESKMLELTNRLQESQEEIQIMIKEK-EEM 1595
```

```
Qy        462 KRIYSSL 468
                ||:  :|
Db       1596 KRVQEAL 1602
```

Search completed: January 2, 2004, 11:00:01
Job time : 20 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 2, 2004, 10:59:27 ; Search time 40 Seconds
(without alignments)
3019.215 Million cell updates/sec

Title: ABU66614
Perfect score: 2399
Sequence: 1 MGFLGTGTWILVLVLPIQAF.....EKGILDKEEAEAIKRIYSSL 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2062	86.0	466	11	Q8R1D7	Q8r1d7 mus musculu
2	1216	50.7	457	13	Q91847	Q91847 xenopus lae
3	198	8.3	7210	5	Q9V7G8	Q9v7g8 drosophila
4	198	8.3	9270	5	Q8MLD9	Q8mld9 drosophila
5	197	8.2	2081	10	Q9LH98	Q9lh98 arabidopsis
6	196.5	8.2	1175	16	Q8XNW6	Q8xnw6 clostridium
7	196	8.2	2139	5	Q07569	Q07569 entamoeba h
8	191.5	8.0	2760	5	Q8I5Y2	Q8i5y2 plasmodium
9	189	7.9	1434	5	Q8I492	Q8i492 plasmodium
10	188.5	7.9	1661	5	Q06166	Q06166 plasmodium
11	184	7.7	951	5	O96229	O96229 plasmodium
12	183.5	7.6	1790	3	Q07380	Q07380 saccharomyc
13	183.5	7.6	2612	5	Q8I5X5	Q8i5x5 plasmodium
14	180.5	7.5	3504	5	Q8IL45	Q8il45 plasmodium
15	179.5	7.5	1320	11	Q9JK25	Q9jk25 rattus norv
16	179.5	7.5	1510	5	Q25920	Q25920 plasmodium
17	179	7.5	853	10	Q9LFE4	Q9lfe4 arabidopsis
18	179	7.5	1558	5	O96275	O96275 plasmodium
19	179	7.5	1786	5	Q9U0P0	Q9u0p0 plasmodium
20	179	7.5	2269	5	Q26223	Q26223 plasmodium
21	179	7.5	2747	5	Q9BJX9	Q9bjx9 plasmodium
22	178.5	7.4	1323	5	Q9NB35	Q9nb35 plasmodium
23	178.5	7.4	3317	16	Q8EWP8	Q8ewp8 mycoplasma
24	178	7.4	840	5	Q8I5X4	Q8i5x4 plasmodium
25	177.5	7.4	1464	5	Q8IIF6	Q8iif6 plasmodium
26	177	7.4	1112	16	Q8EWQ1	Q8ewq1 mycoplasma
27	177	7.4	1871	10	Q9SRD5	Q9srd5 arabidopsis
28	176.5	7.4	682	5	Q8IBV6	Q8ibv6 plasmodium
29	175.5	7.3	1964	5	Q8SWQ7	Q8swq7 loligo peal
30	175	7.3	10578	5	Q8ISF5	Q8isf5 caenorhabdi
31	175	7.3	18519	5	Q8ISF6	Q8isf6 caenorhabdi
32	175	7.3	18534	5	Q8ISF7	Q8isf7 caenorhabdi
33	174.5	7.3	924	5	O15738	O15738 dictyosteli
34	174.5	7.3	1263	5	Q8IK49	Q8ik49 plasmodium
35	173.5	7.2	1795	5	Q8IE35	Q8ie35 plasmodium
36	172.5	7.2	1134	2	Q99051	Q99051 streptococc
37	172.5	7.2	1979	5	O96133	O96133 plasmodium
38	172	7.2	495	5	Q8IEK6	Q8iek6 plasmodium
39	171.5	7.1	2083	5	Q9N435	Q9n435 caenorhabdi
40	171.5	7.1	3484	5	P91257	P91257 caenorhabdi
41	171.5	7.1	5507	5	Q8IHN3	Q8ihn3 plasmodium
42	170.5	7.1	2771	5	Q26216	Q26216 plasmodium
43	170	7.1	1140	4	Q9ULE5	Q9ule5 homo sapien
44	170	7.1	1804	11	Q9Z1Q1	Q9z1q1 mesocricetu
45	170	7.1	3130	5	Q9BK46	Q9bk46 plasmodium

ALIGNMENTS

RESULT 1
Q8R1D7
ID Q8R1D7 PRELIMINARY; PRT; 466 AA.
AC Q8R1D7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to secretogranin III.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC024785; AAH24785.1; -.
SQ SEQUENCE 466 AA; 52783 MW; AB1CB9667276315C CRC64;

Query Match 86.0%; Score 2062; DB 11; Length 466;
Best Local Similarity 87.0%; Pred. No. 2.1e-94;
Matches 410; Conservative 17; Mismatches 36; Indels 8; Gaps 2;

```
Qy          1 MGFLGTGTWILVLVL---PIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKT 57
              |||| ||:|||||||   |||||||| ||||||||||||||||||||||||||| |||||
Db          1 MGFLWTGSWILVLVLNSGPIQAFPKPEGSQDKSLHNRELSAERPLNEQIAEAEADKIKKA 60

Qy         58 YPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLI 117
              :| |:|| :|||| ||||||:|||||| :|||||||| | ||:||||| |||||||||
Db         61 FPSESKPSESNYSSVDNLNLLRAITEKETVEKERQSIRSPPFDNQLNVEDADSTKNRKLI 120

Qy        118 DDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLN 177
              |:|||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db        121 DEYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIATRIYEENDRGVFDKIVSKLLN 180

Qy        178 LGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
              |||||||||||||||| |||||||||||||  :|||| |||| |||||||||:||:|
Db        181 LGLITESQAHTLEDEVAEALQKLISKEANNYEETLDKPTSRTENQDGKIPEKVTPVAAVQ 240
```

```
Qy        238 DGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              ||    |||||||||||:|||||||  : ||:|||||||||||||| ||||||||||||
Db        241 DGFTNRENDETVSNTLTLSNGLERRTNPHREDDFEELQYFPNFYALLTSIDSEKEAKEKE 300

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              |||||||||||||||||||||||||||||||||||| |||||||||||| ||:
Db        301 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDETIALQTKNKLEKNTTDS-----KT 355

Qy        358 PSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIE 417
              | ||| |||||||||||||||||||||||||||||| ||||||  ||||||||||||||||
Db        356 PPEKSQEETDSTKEEAAKMEKEYGSLKDSTKDDNSNLGGKTDEATGKTEAYLEAIRKNIE 415

Qy        418 WLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              |||||:||||||||||||||||:|||||||||||||||| |||||||||
Db        416 WLKKHNKKGNKEDYDLSKMRDFINQQADAYVEKGILDKEEANAIKRIYSSL 466


RESULT 2
Q91847
ID   Q91847      PRELIMINARY;      PRT;   457 AA.
AC   Q91847;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Secretogranin III.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Int. pituitary;
RX   MEDLINE=96217355; PubMed=8632145;
RA   Holthuis J.C.M., Martens G.J.M.;
RT   "The neuroendocrine proteins secretogranin II and III are regionally
RT   conserved and coordinately expressed with proopimelanocortin in
RT   Xenopus intermediate pituitary.";
RL   J. Neurochem. 66:2248-2256(1996).
DR   EMBL; X92872; CAA63478.1; -.
SQ   SEQUENCE   457 AA;  51903 MW;  B3096E12D4B076D7 CRC64;

  Query Match           50.7%;  Score 1216;  DB 13;  Length 457;
  Best Local Similarity 57.0%;  Pred. No. 1.2e-52;
  Matches  269;  Conservative   63;  Mismatches  106;  Indels   34;  Gaps   10;

Qy          8 TWILVLVLPIQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQS 67
              | |: : : :| || |   ||| :|||||| |||| ||||||  | : :    ||:
Db          9 TAIIGVFVHVQGFPNPSIKQDKDVHNRELSEERPLEEQIAEA--DTVSRGTSTENQEMLK 66

Qy         68 NYSFVDNLNLLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGL 127
              | || |:|  | :: ||:   |   | | |   |:  ::| ||||| || : :|:|||
Db         67 NDSFADDLGFLMSMGEKDGDVKRSGSTRGSFGSNRGTLDDADSTKNLKLAEYFDTTKSMP 126

Qy        128 DHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQAH 187
              |:|::|||||||||  | |||||| ||| |||||||| ||||||||||||||||||:
Db        127 DYKYEDDPDGLHQLD-APFTAEDIVRKIATRIYEENDRGVFDKIVSKLLNLGLITESQAY 185

Qy        188 TLEDEVAEVLQKLISKEANNYEE--------DPNKPTSWTENQAGKIPEKVTPMAAIQDG 239
              ||||||||||:||: || | |:        |       : : ||:           :
Db        186 TLEDEVAEVLQQLIANEAKNREKEAGDLDYSDVRSDVYDDDKREGKM----------ET 234

Qy        240 LAKGENDET--VSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKE 297
              | | |: |:    | :|:  :||| :   ||   :|||||||  |:||::|| : |||:
Db        235 LNKNEDRESSETKNEDSLST-VERRNELSPED---DLQYFPNFNRLVKSLNSENDVKEKK 290

Qy        298 TLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPA 357
              ||:||||||||||||||||:|||||:||||||:|||:| |||| ||   : : |
Db        291 TLLTIMKTLIDFVKMMVKYGTITPEEGVNYLENLDDMIAVQVKNKLGKNFGSHRNAL--- 347

Qy        358 PSE-KSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNI 416
              ||| |: :|:|||||||||||||  |||||| : ::      :|||| | |||||||||
Db        348 PSEDKNADESDSTKEEAAKMEKEYELLKDSTKPEPND--AAESKPKGKAETYLEAIRKNI 405

Qy        417 EWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
              ||||:|:|: || | ||||:: |:|:|||:||||||||||  ||||||||
Db        406 EWLKEHNKEANKGDNDLSKLKQFVNQQADSYVEKGILDKEEANVIKRIYSSL 457


RESULT 3
Q9V7G8
ID   Q9V7G8      PRELIMINARY;      PRT;  7210 AA.
AC   Q9V7G8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG18255 protein.
GN   STRN-MLCK OR CG8304 OR CG18255.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
```

```
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAF58087.2; -.
DR   FlyBase; FBgn0013988; Strn-Mlck.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 13.
DR   SMART; SM00409; IG; 21.
DR   SMART; SM00408; IGc2; 20.
DR   PROSITE; PS50835; IG_LIKE; 20.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   7210 AA;  808965 MW;  6B2E7395C07140D2 CRC64;

  Query Match             8.3%;  Score 198;  DB 5;  Length 7210;
  Best Local Similarity   23.8%;  Pred. No. 0.55;
  Matches  110;  Conservative   83;  Mismatches  180;  Indels   90;  Gaps   19;

Qy         35 ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSI 94
              |: :|:  :|:  |::: ::| :   | || ::         | |    |:||:| :::
Db       4757 EIVSEKITDEKAQESQKKEVKGS---EAKPKKAKV-------LEKKSIEEEKLEDKKEKQ 4806

Qy         95 RSSPLDNKLNVEDVDSTKNRKLIDD--YDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIV 152
               | :| |    :|    : |: |:   :| |  :    :  |      |:   :  | :
Db       4807 TESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDS-EAKPKKAKVLEKKSIEEEKLE 4865

Qy        153 HK----IAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQ-KLISKEANN 207
              :|      : | |:: :|   :|||: :      |||   ::|  |:  : |:: |::
Db       4866 NKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIE 4925

Qy        208 YEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYS 267
               |:  :|    ||          :|| :   | |  ||||  :|     : : :
Db       4926 EEKLEDKKEKQTE-------------SAIDEKFQKAEVSETVSEKIT-----DEKAEESR 4967

Qy        268 EDNFEELQYFPNFYALL--KSIDSEK--EAKEKET------------LITIMKTLID-- 308
              ::  :: :  |    :|  |||: ||  : |||:|              |: : : |
Db       4968 KEEVKDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESAIDEKSQKAEVSETVSEKITDEK 5027

Qy        309 ---FVKMMVKYGTISP------EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPS 359
                 |  ||      |       |:    :| |||     ||: |:   |||  |:
Db       5028 AQESQKKEVKDSEAKPKKAKILEKKSIEIEKLDEKKEKQTETKV---ATDTKSQTVEVSE 5084

Qy        360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWL 419
               : ::   ||: |:|     ||||             | | |    ||       | |
Db       5085 IVLEKISEEKAEESQKVE-----LKDS-------------EAKSKKAKVLEKKSTLKEKL 5126

Qy        420 KKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIK 462
                ::||| |||   :| :    : ||   ||  : :|:   ||
Db       5127 DENDKK-QKEDGATNKSQK--AEAADVVPEK--ISEEKVAEIK 5164


RESULT 4
Q8MLD9
ID   Q8MLD9      PRELIMINARY;      PRT;  9270 AA.
AC   Q8MLD9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG18255-PA.
GN   STRN-MLCK OR CG8304 OR CG18255.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
```

```
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAM70936.1; -.
DR   FlyBase; FBgn0013988; Strn-Mlck.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 29.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00408; IGc2; 29.
DR   SMART; SM00406; IGv; 4.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 26.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Immunoglobulin domain; Transferase.
SQ   SEQUENCE   9270 AA;  1036749 MW;  0DD82D040B27C2AD CRC64;

  Query Match             8.3%;  Score 198;  DB 5;  Length 9270;
  Best Local Similarity   23.8%;  Pred. No. 0.73;
  Matches  110;  Conservative   83;  Mismatches  180;  Indels   90;  Gaps   19;

Qy        35 ELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEKEKIEKERQSI 94
Db      4757 EIVSEKITDEKAQESQKKEVKGS---EAKPKKAKV-------LEKKSIEEEKLEDKKEKQ 4806

Qy        95 RSSPLDNKLNVEDVDSTKNRKLIDD--YDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIV 152
Db      4807 TESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDS-EAKPKKAKVLEKKSIEEEKLE 4865

Qy       153 HK----IAARIYEENDRAVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQ-KLISKEANN 207
Db      4866 NKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIE 4925

Qy       208 YEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTLTNGLERRTKTYS 267
Db      4926 EEKLEDKKEKQTE-------------SAIDEKFQKAEVSETVSEKIT-----DEKAEESR 4967

Qy       268 EDNFEELQYFPNFYALL--KSIDSEK--EAKEKET------------LITIMKTLID-- 308
Db      4968 KEEVKDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESAIDEKSQKAEVSETVSEKITDEK 5027
```

```
Qy       309 ---FVKMMVKYGTISP------EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPS 359
Db      5028 AQESQKKEVKDSEAKPKKAKILEKKSIEIEKLDEKKEKQTETKV---ATDTKSQTVEVSE 5084

Qy       360 EKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWL 419
Db      5085 IVLEKISEEKAEESQKVE-----LKDS------------EAKSKKAKVLEKKSTLKEKL 5126

Qy       420 KKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIK 462
Db      5127 DENDKK-QKEDGATNKSQK--AEAADVVPEK--ISEEKVAEIK 5164


RESULT 5
Q9LH98
ID   Q9LH98      PRELIMINARY;      PRT;  2081 AA.
AC   Q9LH98;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Genomic DNA, chromosome 3, BAC clone: T19N8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP002057; BAB03174.1; -.
SQ   SEQUENCE   2081 AA;  232851 MW;  D3603E1F85EFFF29 CRC64;

  Query Match             8.2%;  Score 197;  DB 10;  Length 2081;
  Best Local Similarity   25.2%;  Pred. No. 0.14;
  Matches  120;  Conservative   68;  Mismatches  175;  Indels  114;  Gaps   21;

Qy        27 QDKSLHNRELSAERPLNEQIAEAEED----------------KIKKTYPPENKPGQSNYS 70
Db       707 EDKKLENKESQTDSKDDKSVDDKQEEAQIYGGESKDDKSVEAKGKKKESKENKKTKTNEN 766

Qy        71 FVDNL--NLLKAITEKEKIEK--ERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSG 126
Db       767 RVRNKEENVQGNKKESEKVEKGEKKESKDAKSVETKDN-KKLSSTENR----DEAKERSG 821

Qy       127 LDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEEN---DRAVFDKIVSKLLNLGLITE 183
Db       822 EDNKEDKEESKDYQ-------------SVEAKEKNENGGVDTNVGNKEDSKDLKDDRSVE 868

Qy       184 SQAHTLED-----EVAEVLQKLISKE----ANNYEEDPNKPTSWTENQAGKIPEKVTPMA 234
Db       869 VKANKEESMKKKREEVQRNDKSSTKEVRDFANNMDIDVQKGSG--ESVKYKKDEK----- 921

Qy       235 AIQDGLAKGENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAK 294
Db       922 --KEG-NKEENKDTI-NTSSKQKGKDKKKKKKESKN--------------SNMKKKEED 962

Qy       295 EKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKL 354
Db       963 KKEYVNNELKKQEDNKKETTKSENSKLKE-----ENKD--------NKEKKESEDSASK- 1008

Qy       355 FPAPSEKSHEETDS-TKEEAAKMEKEYGSLKDSTKDDNSNPG-------------GKTDE 400
```

```
Db       1009 --NREKKEYEEKKSKTKEEAKKEKKKSQDKKREEKDSEERKSKKEKEESRDLKAKKKEEE 1066

Qy        401 PKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEE 457
Db       1067 TKEKKES------ENHKSKKKEDKKEHEDNKSMKKEED--KKEKKKHEESKSRKKEE 1115


RESULT 6
Q8XNW6
ID   Q8XNW6      PRELIMINARY;      PRT;  1175 AA.
AC   Q8XNW6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Probable exonuclease.
GN   SBCC OR CPE0216.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003185; BAB79922.1; -.
DR   InterPro; IPR003439; ABC_transporter.
KW   Exonuclease; Complete proteome.
SQ   SEQUENCE   1175 AA;  136878 MW;  7C6D2366525019C1 CRC64;

  Query Match             8.2%;  Score 196.5;  DB 16;  Length 1175;
  Best Local Similarity   22.6%;  Pred. No. 0.078;
  Matches  116;  Conservative   83;  Mismatches  196;  Indels  119;  Gaps   21;

Qy         29 KSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNL-NLLKAITEKE-- 85
Db        256 KELYDKRIEEE----SLVSRSEEIKSFKERVEISNKADKVIVFINNLEEILKEINKEDLK 311

Qy         86 --KIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDG 143
Db        312 FSELNKKLEELINLREENKLKFEEFTKKKEEKLPDLRLKKEKLLES--QKERDILFQIKA 369

Qy        144 TPLTAEDIVHKI---------AARIYEENDRAVFDKIVSK---------------LLNLG 179
Db        370 DGVKLKEACKKIFEDRSKCDTKLNSIEENEKRLNEELKEKEERKEELFVHEEFKNKINSG 429

Qy        180 LITESQAHTLEDEVAEVLQKLI--SKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQ 237
Db        430 LFILNSYESLDKQFNEIKSEEVELKKYIKNLTEDKEKSEK-------DLKVKVESLSKIR 482

Qy        238 DGLAK------------------GENDETVSNTLTLTNGLERRTKTYSEDNFEE-LQYF 277
Db        483 DKLESLLKETPGDSNSILEKQIKLGEYREKLNKYKEIKNSLEESLKT--KNNFEEKLKTF 540

Qy        278 PNFYALLKSIDSEKEAKEKETLITIMKT--LIDFVKMMVKYGTISPEEGVSYLE------ 329
Db        541 ENQKLLL-----EKEVRELKDYINKVKVEELAHKLRENLVEGECCPVCGSTHHELNKVEK 595

Qy        330 -NLDE----MIALQTKNKLEKNATDNISKL---FPAPSEK------SHEETDSTKEEAAK 375
Db        596 INLEESNEKTILLESKEEKLKELILEFSKIEATLEYENKKIEELNISIEEVGEVNEERLK 655

Qy        376 -MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLS 434
Db        656 FLEEEFNTLKDKIEEFNLKK--------------ENLEKDLEKLK--EEKNNLE----- 693

Qy        435 KMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL 468
Db        694 ---NIFNKAEVILCEKIVREKEIASKIKELDKEL 724
```

```
RESULT 7
Q07569
ID   Q07569      PRELIMINARY;      PRT;  2139 AA.
AC   Q07569; O02504;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Myosin heavy chain.
GN   MHCA.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HMI:IMSS;
RX   MEDLINE=93295430; PubMed=8515774;
RA   Raymond-Denise A., Sansonetti P., Guillen N.;
RT   "Identification and characterization of a myosin heavy chain gene
RT   (mhcA) from the human parasitic pathogen Entamoeba histolytica.";
RL   Mol. Biochem. Parasitol. 59:123-131(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HMI:IMSS;
RA   Guillen N.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L03534; AAB48065.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00242; MYSc; 1.
SQ   SEQUENCE   2139 AA;  245225 MW;  C68307341DB51DD1 CRC64;

  Query Match             8.2%;  Score 196;  DB 5;  Length 2139;
  Best Local Similarity   21.0%;  Pred. No. 0.17;
  Matches  107;  Conservative  108;  Mismatches  191;  Indels  104;  Gaps   23;

Qy         20 FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPEN-------KPGQSNYSFV 72
Db        859 FEEEGKKKDKEIED--------LKKKLAE----EIKKREAAENALASATAKTGELEAKIQ 906

Qy         73 D----------------------NLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVD- 109
Db        907 DLEDKISELESKLSAAELDKQELNLKIENLEEDKEELKETIDNLKGDLKDSKLKGEDLEV 966

Qy        110 ---------STKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIY 160
Db        967 EITELNSQINTLN-ATVDKKDKTIAEMQESIDEKEDEITKLKGDIKLLEEEKDDL----- 1020

Qy        161 EENDR----AVFDKIVSKLLNLGLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPT 216
Db       1021 -EQDRADVSATKDDIAKKLNKITI----ECEDAKDEIAKLEQELEDEE--NKNKDLTNEL 1073

Qy        217 SWTENQAGKIPEKVTPMAAI-------QDGLAKG-ENDETVSNTLTLTNG-LERRTKTYS 267
Db       1074 QQTQLKLGETEKSLAAQVAATKKASDERDTLSQNLENEKLTTKNLTKTKADLEKKISGLK 1133

Qy        268 EDNFEELQYFPNFYALLKSID-----SEKEAKEKETLITIMKTLIDFV-KMMVKYGTISP 321
Db       1134 QD-YEDLEDDKN------KIEGDLRNAQRKIKELDDEITKGADVSQYLQKQKEEY----- 1181

Qy        322 EEGVSYLENLDEMIALQTKNKLEKNATDNISKLFPAPSEKSHEETDSTKEEAAK----ME 377
Db       1182 ESQIAKMQEEKEAIGNDVKNK-EKTIKE--KELEIQSLQEKLDETEVEKEDAEKKKKEIE 1238

Qy        378 KEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL-SKM 436
Db       1239 KEMKALQEE-KENVESSKNSTEKDKKKLEDNLKDTQKKLDDMTADNEKLKAKAKDLEAQL 1297
```

```
Qy        437 RDFINKQADAYVEKGILDKEEAEAIKRIYS 466
              :  :    |  :  :|:|::|:: | : |
Db       1298 NEVQDNHEKAVADAELLNKKKAQSDKELNS 1327
```

RESULT 8
Q8I5Y2
ID Q8I5Y2 PRELIMINARY; PRT; 2760 AA.
AC Q8I5Y2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL0315C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014845; AAN36152.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2760 AA; 330896 MW; EE9964C4845181AC CRC64;

Query Match 8.0%; Score 191.5; DB 5; Length 2760;
Best Local Similarity 21.1%; Pred. No. 0.37;
Matches 121; Conservative 92; Mismatches 207; Indels 153; Gaps 24;

```
Qy         29 KSLHNRELSAERPLNEQIAEAEE--------------DKIKKTYPPENKPGQSNYSFVD 73
              |:    |   |: :|:|:   :|                ::|| |   | : : |
Db       1452 KNEAKEEEETEKNINDQMVNEKEMDVNSKNEREIVQVHNEIKNTNNKEEEGKKKNLLKEK 1511

Qy         74 NLN--LLKAITEKEKIEKERQ--SIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDH 129
              :|  |   | :::| ||::  ::   |: : |  :::   : | :   |     |:
Db       1512 EINDCLNDYINKQKKKEKKKNNWAMYGRPIVKRQNNRNINIKNDLKKLYSSKSESGFNDY 1571

Qy        130 KF-----------QDDPDGLHQLDGTPLTAE---DIVH---------KIAARIYEEN-- 163
              |           :|| |  :|      |   ||:|          |:   ||| :
Db       1572 AFYAERFFEVITGYNSEPDYLSDIDNQAKNEENKNDIIHNNNIIKISKKMKENIYENSPF 1631

Qy        164 ---DRAVFD-------------------------KIVSKLLNLGLITESQA-------- 186
                | :::                          | ::| :::   |:  :
Db       1632 HTYGRPIYEKKSKNPNNYNKIKSTTHNAILKKKRKKTLNKSISINSFTKMNSSNNKIVKR 1691

Qy        187 -----HTLEDEVAEVLQKLISKEANNYEEDP----------------NKPTSWTENQAGK 225
                 :|:::     ::|:| || |   ||                  ::      |: :  |
Db       1692 TSIKNNTIDNYNNSTIKKIIHKEQN--VEDQGYIDLKTKRKLIYDALDEINQQTQQKNLK 1749

Qy        226 IPEKVTPMAAIQDGLAKGENDETVSNTLTLTNG--LERRTKTYSEDNFEELQYFPNFYAL 283
              | | :|     | |   |:|   ||| :  |    :|: |:    || || |:     ||
Db       1750 ITENIT-----QVGKKHGQN---VSNIIKNTGAMLIEKITRK-GNDNNEEDQFSEELKAL 1800

Qy        284 LKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLENLDEM--------I 335
              | :   || |  | | |: :   :  | : :  ::  ||      :|::||        |
Db       1801 EK-LKKLKELKRIEELKTLEE---EEKKRLEELNSLKVEEEKKKKKNMEEMKKMEEMKKI 1856

Qy        336 ALQTKNKLEK-----NATDNISKL--------FPAPSEKSHEETDSTKEEAAKM--EKEY 380
              || | : ||        : | ||             | :  |    :||  ||  |||
Db       1857 ELQRKKEEEKLLQYSRRQEQIRKLEELKKKEELKLIEEMNKLEEQRKREEIKKMEEEKER 1916
```

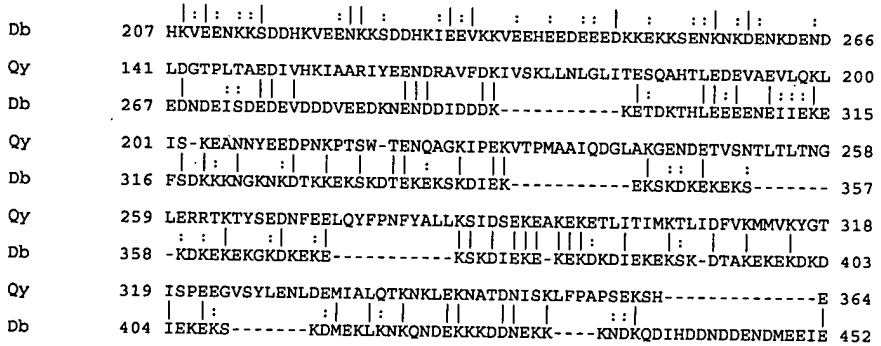
```
Qy        381 GSLKDSTKDDNSNPGGKTDEPKGKTEAYL-------EAIRKNIE--WLKKHDKKGNKEDY 431
              ||   :          | :| : | :  |       |  || :|  ||::  :|   :
Db       1917 EKLKKIKEQQKKEEMKKQEEQRKKEQIKLMEEMRKFEEQRKKVELKWLEEEKRKEELKKL 1976

Qy        432 DLSKMRDFINKQADAYVEKGILDKEEAEAIKRI 464
              :  | |:   |: :   :|  :|:  | ||::
Db       1977 EEQKKRE-ERKKMEEEEKKREEEKKREEEIKKM 2008
```

RESULT 9
Q8I492
ID Q8I492 PRELIMINARY; PRT; 1434 AA.
AC Q8I492;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mature parasite-infected erythrocyte surface antigen (MESA) or
DE PfEMP2.
GN MESA.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51374.1; -.
SQ SEQUENCE 1434 AA; 168287 MW; AB0005F9DC26C989 CRC64;

Query Match 7.9%; Score 189; DB 5; Length 1434;
Best Local Similarity 21.1%; Pred. No. 0.23;
Matches 101; Conservative 96; Mismatches 180; Indels 102; Gaps 23;

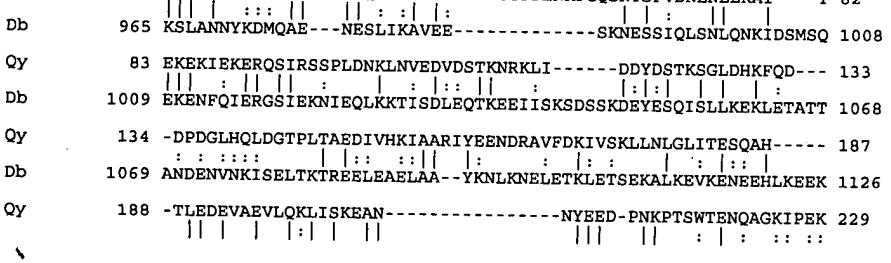
```
Qy         27 QDKSLHNRELSAERPLNEQ-----IAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI 81
              :|| |   :    :  |::     : | ::: :|      |   |: :   : :    | |
Db        501 KDKVLGEGDKEDVKEKNDEQKDKVLGEGDKEDVK-----EKNDGKKD-KVIGSEKTQKEI 554

Qy         82 TEKEKIEKE-----RQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQD--- 133
              |||:||      :: ::    :|  :  |   |  ::|   :  |  :  : :|
Db        555 --KEKVEKRVKKKCKKKVKKGIKEN--DTEGNDKVKGPEII--IEEVKEEIKKQVEDGIK 608

Qy        134 --DPDGLHQLDGTPLTAEDIVHKIAARIYE---ENDRAVFDKIVSKLLNLGLITESQAHT 188
               | :|  :: |  :  |::  :|  :: |   |||    ||:        :|||
Db        609 ENDTEGNDKVKGPEIITEEVKEEIKKQVEEGIKENDTEGNDKVKGP----EIITEEVKEE 664

Qy        189 LEDEVAEVLQ--------KLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGL 240
              :: :| | ::         ||| :|     |        : |||:     :||      | : :
Db        665 IKKQVEEGIKENDTESKDKLIGQEIITEEVKEGIKENDTENK-----DKVIGQEIITEEV 719

Qy        241 AKG--ENDETVSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKET 298
              :|  |||              | : |   ::   |          : | |: ::|   ||
```

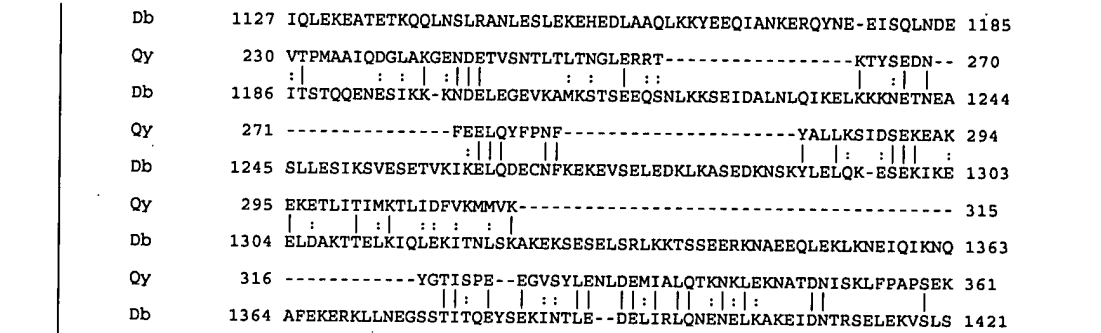
```
Db        720 KEGIKEND------------TENKDKVIGQEIITE--------EVKKEIEKQEEKGNKEN 759

Qy        299 LITI------MKTLIDFVKMMVKYGTISPEEGV--SYLENLD-----EMIALQTKNKLEK 345
              :: |        : :|: || ::|       |:|:  :: |: |      |:|  : | ::||
Db        760 ILEIKDIVIGQEVIIEEVKKVIKK---KVEKGIKENHTESKDKVIGQEIIVEEVKEEIEK 816

Qy        346 NATDNISKLFPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKT 405
                : |             |    :|::    |   |   |  ::   |     | : | |
Db        817 QVEEGIK----------ENDTESKDKVIGQEVIKG---DVNEEGPENKDKVTKQEKVK- 861

Qy        406 EAYLEAIRKNIEWLKKHDKKGNKEDYDLSK--MRDFINKQADAYVEKGILDKEEAEAIK 462
              |   |  :|  : :|| : |  ::|  : |  |:: :|::  |  :| |  ::| | :|
Db        862 EVKKEVKKKVKKRVKKRNNKNERKDNVIGKEIMKEDVNEKDTANKDKEIEQEKEKEEVK 920


RESULT 10
Q06166
ID   Q06166      PRELIMINARY;      PRT;  1661 AA.
AC   Q06166;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Mature PARASITE-infected erythrocyte surface antigen (Antigenic
DE   protein PFEMP2).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PALO ALTO;
RX   MEDLINE=92158014; PubMed=1741020;
RA   Coppel R.L.;
RT   "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT   binds human erythrocyte protein 4.1.";
RL   Mol. Biochem. Parasitol. 50:335-347(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PALO ALTO;
RA   Kun J.F.J., Waller K.L., Coppel R.L.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 797-850 FROM N.A.
RX   MEDLINE=93122844; PubMed=1478701;
RA   Saul A., Yeganeh F., Howard R.J.;
RT   "Conservation of repeating structures in the PfEMP2/MESA protein of
RT   Plasmodium falciparum.";
RL   Immunol. Cell Biol. 70:353-355(1992).
DR   EMBL; AF056936; AAC13303.1; -.
DR   EMBL; S52458; AAB24869.1; -.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Antigen.
SQ   SEQUENCE   1661 AA;  195479 MW;  AF340527D85A9D29 CRC64;

  Query Match             7.9%;  Score 188.5;  DB 5;  Length 1661;
  Best Local Similarity   19.9%;  Pred. No. 0.29;
  Matches   94;  Conservative   90;  Mismatches  189;  Indels   99;  Gaps   15;

Qy         22 KPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAI 81
              |   ::|| :   :   |    |:: | || | |:    : :                 :
Db        895 KDTANKDKEIEQEKEKEEVKEKEEVKEKEEVKEKEEVKEKEE----------------V 937

Qy         82 TEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQL 141
               |||:::: :      :  |  |:: |:    | |:  :  |  :    : | :   ::
Db        938 KEKEEVKEKEEVKEKEEVKEKEEVKEKDTESKDKEIEQ-EKEKEEVKEVKEKDTENKDKV 996

Qy        142 DGTPLTAEDIVHKIAARIYEEN------DRAVFDKIVSKLLNLGLITESQAHTLEDEVAE 195
                |  :  |:|  ::  |: : |      |  :  :|::::|     : :    | | |
Db        997 IGQEIIIEEIKKEVKKRVKKRNNKNENKDNVIVQEIMNEMLRKDTANKDKVIEQEKEKEE 1056

Qy        196 VLQKLISKEANNYEEDPNKPTSWTENQAGKIPEKVTPMAAIQDGLAKGENDETVSNTLTL 255
                | :|   ||    :|                |:|             | :|
Db       1057 VKEKEEVKEKEEVKE----------------KEEVK-----------EKEEV------- 1081

Qy        256 TNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKET----LITIMKTLIDFVK 311
                    :   | |:|   | :          |  :  || |||:|     :    : :|: :|
Db       1082 -----KEKDTESKDKEIEQE---------KEKEEVKEVKEKDTENKDKVIGQEIIIEEIK 1127

Qy        312 MMVKYGTISPEEGVSYLENLDEMIALQTKNK-LEKNATDNISKLFPAPSEKSH-EETDST 369
                ||      ::  :  || | :|  :  |: : :  | |  |:     ||   :| :
Db       1128 KEVKKRV---KKRNNKNENKDNVIVQEIMNEDVNEKDTANKDKVIEQEKEKEEVKEKEEV 1184

Qy        370 KEEAAKMEKEYGSLKDS-----------TKDDNSNPGGKTDEPKGKT---EAYLEAIRKN 415
                ||:    |||    ||:           |::      |  | | |    |  :| |:|
Db       1185 KEKEEVKEKEEVKEKDTESKDKEIEQEKEKEEVKEVKEKDTENKDKVIGQEIIIEEIKKE 1244

Qy        416 IE--WLKKHDKKGNKEDYDLSK-MRDFINKQADAYVEKGILDKEEAEAIKRI 464
              ::    |:::|  ||::  : : | : :|::  |  :| |  ::| | :| :
Db       1245 VKKRVKKRNNKNENKDNVIVQEIMNEDVNEKDTANKDKVIEQEKEKEEVKEV 1296


RESULT 11
O96229
ID   O96229      PRELIMINARY;      PRT;  951 AA.
AC   O96229;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFB0680W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=99021743; PubMed=9804551;
RA   Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA   Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA   Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA   Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA   Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE001410; AAC71925.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   951 AA;  112486 MW;  AC8D889358A84F4F CRC64;

  Query Match             7.7%;  Score 184;  DB 5;  Length 951;
  Best Local Similarity   22.0%;  Pred. No. 0.25;
  Matches  104;  Conservative   76;  Mismatches  182;  Indels  110;  Gaps   17;

Qy         25 GSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKAITEK 84
              | || |  | |   :: : | : | || | ::    ::|  ::  |            :
Db        161 GKQDISNSNAE--NKKDVKEGVKELEEKKKEEKISDDHKVEENKKS------------DD 206

Qy         85 EKIEKERQSIRSSPLDNKLN----VEDVDSTKNRKLIDDYDSTKSGLDHKFQDDPDGLHQ 140
```



RESULT 12
Q07380
ID Q07380 PRELIMINARY; PRT; 1790 AA.
AC Q07380; P89892;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 206.5 kDa protein YDL058W.
GN USO1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z74106; CAA98621.1; -.
DR EMBL; Z74105; CAA98620.1; -.
DR SGD; S0002216; USO1.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Uso1_p115_C.
DR InterPro; IPR006953; Uso1_p115_head.
DR Pfam; PF04871; Uso1_p115_C; 1.
DR Pfam; PF04869; Uso1_p115_head; 1.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;

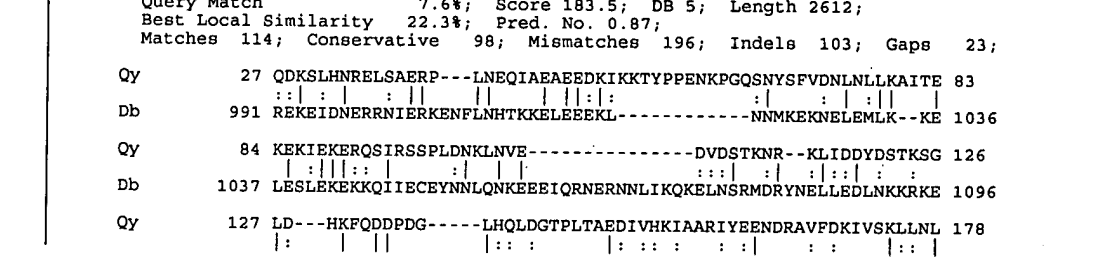
Query Match 7.6%; Score 183.5; DB 3; Length 1790;
Best Local Similarity 20.3%; Pred. No. 0.56;
Matches 120; Conservative 91; Mismatches 200; Indels 179; Gaps 23;





RESULT 13
Q8I5X5
ID Q8I5X5 PRELIMINARY; PRT; 2612 AA.
AC Q8I5X5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL0350C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014845; AAN36159.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2612 AA; 311421 MW; C950B6E466E3D1C5 CRC64;

Query Match 7.6%; Score 183.5; DB 5; Length 2612;
Best Local Similarity 22.3%; Pred. No. 0.87;
Matches 114; Conservative 98; Mismatches 196; Indels 103; Gaps 23;



```
Db       1097 LEMEKMKLLDDVQEERIKFLNEKNNMQKEKENEINYMKEELRKERILMIEEVEKMKVIML 1156

Qy        179 GLI----------TESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGKIPE 228

Db       1157 EDIEKNKEKMIKNVEKENEKLKDEIEKERRNMI----QNLEEEKKEFKLYLEQKYKENFE 1212

Qy        229 KVTPMAAIQDGLAK--GENDETVSNTLTLTNGLERRTKTYSEDNFEELQ--YFPNF---- 280

Db       1213 NE------KSGLAKKFDEENEKLQNEI----GNEKRKLHKERDNFEEQKKIYEEEFRNKC 1262

Qy        281 ----YALLKSID--SEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYL------ 328

Db       1263 EKYEESIQKKYDVLEEDKNKMKYLIIKEQEELENYKKNI--YLDIEEEKEKLYVQQEKLN 1320

Qy        329 ---ENL---DEMIALQTKN-KLEKNATDNISKL----FPAPSEKSHEETDSTKEEAAKME 377

Db       1321 LEKENLLVEKEQIEIELKNYKNFKEKEENDIKIKIINLSQQQEDLNKEKENIEKEKEQIE 1380

Qy        378 KEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKK-GNKEDYDLSKM 436

Db       1381 KRKYDL-DEREESLNNDKAQMEESRKIFDEQLDKIKKNKEELLNYDKELKNREIHLMEKE 1439

Qy        437 RDFINKQADAYVEKGILDKEEAEAIKRIYSS 467

Db       1440 KEYNKKLEESNKKKESLNSFDVEL--KEYSS 1468
```

RESULT 14
Q8IL45
ID Q8IL45 PRELIMINARY; PRT; 3504 AA.
AC Q8IL45;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0404.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014822; AAN37017.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3504 AA; 408303 MW; B8454D48D55BE4F0 CRC64;

Query Match 7.5%; Score 180.5; DB 5; Length 3504;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 104; Conservative 93; Mismatches 172; Indels 139; Gaps 22;

```
Qy         32 HN----RELSAE-------RPLNEQIAEAE-----EDKIKKTYPPENKPGQ--------- 66

Db        982 HNLIGKNEVSDETMNEYIISPKSEIINEQENTPQYEEVLSKNINNDKDNGKRINNNDQLE 1041

Qy         67 ----SNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDN----KLNVEDVDSTKNRKLID 118

Db       1042 SPILSNEKKTDDIHI---IEEKEKIDK---PIHNNDLENQNHEKKHIQD-------DKLE 1088

Qy        119 DYDSTKSGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNL 178

Db       1089 EYSTLPNSKEMK--------------NISGDNIENNLLSNKKEENHKIIHENIEEDTTSK 1134

Qy        179 GLITESQAHTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTEN-----QAGKI------- 226

Db       1135 DNEKNMQNYIIHDDLG---KKDISRENTGEDISPNKYLKNKENYGILDDKGKIKIYEENN 1191

Qy        227 PEKVTPMAAIQDGLAKGENDETV----------------SNTLTLTNGLERRTKTYSEDN 270

Db       1192 EEYVTDVGSGENKVSEGNSNDNIMPEEYGKNKNSQENSEDNIMTEKNGKNKSSEENTDDN 1251

Qy        271 FEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFVKMMVKYGTISPEEGVSYLEN 330

Db       1252 ----------------IMTEKNGKDKSSEENANDNI-----MTEKNGKNKSSQG-----N 1285

Qy        331 LDEMIALQ---TKNKLEKNATDNISKLFPAPSEKSHEETDST-------KEEAAK----- 375

Db       1286 IDDNIMTEKNGKNNNSQENSDDNIMTEEYEKNKNSQENTDDNIMTEKNGKNKSSEENTDD 1345

Qy        376 --MEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDL 433

Db       1346 NIMTEKNG--KDKSSEENANDNVMT-EKNGKNKSSEENADDNIMTEKNGKNKSSEENADD 1402

Qy        434 SKMRD--FINKQADAYVEKGILDKEEAE 459

Db       1403 NIIREENGKNKSSQENSDDNIMTEEYGE 1430
```

RESULT 15
Q9JK25
ID Q9JK25 PRELIMINARY; PRT; 1320 AA.
AC Q9JK25;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CLIP-170.
GN CYLN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RA Galjart N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RX MEDLINE=21185938; PubMed=11290329;
RA Akhmanova A., Hoogenraad C.C., Drabek K., Stepanova T., Dortland B.,
RA Verkerk T., Vermeulen W., Burgering B.M., de Zeeuw C.I., Grosveld F.,
RA Galjar N.;
RT "CLASPs are CLIP-115 and -170 associating proteins involved in the
RT regional regulation of microtubule dynamics in motile fibroblasts.";
RL Cell 104:923-935(2001).
DR EMBL; AJ237670; CAB92974.1; -.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_GLY; 2.
DR SMART; SM00343; ZnF_C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS50245; CAP_GLY_2; 2.
SQ SEQUENCE 1320 AA; 148264 MW; 87BE0F0463D20E9D CRC64;

Query Match 7.5%; Score 179.5; DB 11; Length 1320;
Best Local Similarity 22.6%; Pred. No. 0.62;
Matches 134; Conservative 77; Mismatches 211; Indels 171; Gaps 27;

```
Qy         22 KPGGSQDKSLH-NRELSAERPLNEQIAEAEEDKIKKTYPPENKPGQSNYSFVDNLNLLKA 80

Db        525 KPPGDVDMSLSLLQEISA---LQEKLEVTHTDHQNEVTSLKDHFGTREEMFQKEIKALHA 581

Qy         81 ITEKEKIEKERQSIRSSPLD--NKLNVEDV---------------------------- 108
```

```
Db        582 AT--EKLSKENESLRSK-LDHANKENSDVIALWKSKLETAIASHQQAMEELKVSFSKGIG 638

Qy        109 -DS-------TKNRKLIDDYDSTKSGLDHKFQDDPDGLHQLDGTPLTA---------EDI 151

Db        639 TDSAEFAELKTQIERLRLDYQHEIESLQSK-QDSERSAHAKEMESMKAKLMKIIKEKEDS 697

Qy        152 VHKIAARIYEENDRAV--FDKIVSKLLNLGLITESQAHTLEDEVA------EVLQKLISK 203

Db        698 LEAVKARLDTAEDQHLVEMEEMLSKLQEAEIKKEKFASASEEAVSTQTSMQDTVNKLHQK 757

Qy        204 EANNYEEDPNKPTSWTENQAGKIPEKVTPMAA-------IQDGLAKG----END--ETV- 249

Db        758 -----EEQFNMLSSELE----KLRENLTDMEAKFKEKDEREDQLVKAKEKLENDIAEIMK 808

Qy        250 -----SNTLTLTNGLERRTKTYSEDNFEELQY-----FPNFYALLKSI--------DSEK 291

Db        809 MSGDNSSQLTKMND-ELRLK---ERSVEELQLKLTKANENASLLQKSIGEVTLKAEQSQQ 864

Qy        292 EA--KEKETLITIMKTLIDFVKMM-----------VKYGTISPEEGVSY---LENLDEMI 335

Db        865 EAAKKHEEEKKELENKLLELEKKMETSHYQCQDLKAKYEKASSETKIKHEEILQNFQKML 924

Qy        336 ------------------------------------ALQTKNKLEKNATDNISKL-- 354

Db        925 VDTEDKLKAAQEANRDLMQDMEELKSQADKAKAAQTAEDAMQIMEQMTKEKTETLASLED 984

Qy        355 FPAPSEKSHEETDSTKEEAAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRK 414

Db        985 TKQTNAKLQSELDTLKENNLKTVEELNKSKELLNEENQ----KMEEFKKEIETLKQAAAQ 1040

Qy        415 NIEWLKKHDKKGNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSS 467

Db       1041 KSQQLSALQEENVKLAEELGRTRDEVTSHQKLEEERSVLNNQLLEMKKSLPSN 1093
```

Search completed: January 2, 2004, 11:01:55
Job time : 44 secs